MAITRISE de MATHÉMATIQUES STATISTIQUE

M. Gradinaru

2002-2005

Avant propos

Ces notes sont une rédaction du cours oral en amphithéâtre. Il s'agit d'un document de travail et pas d'un ouvrage ; il est destiné à la distribution aux étudiants de Maîtrise de Mathématiques de l'Université de Nancy. Certaines parties de ces notes sont inspirées de notes de cours (et je remercie vivement leurs auteurs) rédigées par F. Castell et B. Roynette. Je remercie B. Roynette pour la lecture attentive des formes préliminaires du manuscrit.

Vandœuvre-lès-Nancy, janvier 2002 - mai 2005 M. Gradinaru

Table des matières

Chapitre 1

Estimation des paramètres

1.1 Échantillon

On associe à une expérience aléatoire une variable aléatoire X, définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Sans perte de généralité on regardera l'espace mesurable d'arrivée de X, noté $(E, \mathcal{B}(E))$, muni de la probabilité Q_X loi de X. Typiquement E est $\mathbb{R}$ si X est une variable réelle ou $\mathbb{R}^d$ si X est un vecteur aléatoire d-dimensionnel ; $\mathcal{B}(E)$ est une tribu sur E, par exemple la tribu borélienne de E.

On répète l'expérience dans les mêmes conditions n fois et on observe les valeurs $x_1^{\mathrm{obs}}, \ldots, x_n^{\mathrm{obs}}$. On regardera ces valeurs comme les valeurs des variables indépendantes, de même loi que X, $(X_1, \ldots, X_n)$. Les valeurs possibles de $(X_1, \ldots, X_n)$ seront notées $(x_1, \ldots, x_n)$. On dit que $\mathbf{X} = (X_1, \ldots, X_n)$ est un n-**échantillon** de loi Q_X. Il s'agit d'un vecteur aléatoire à valeurs dans E^nde loi donnée par

$$\mathbb{P}(\mathbf{X} \in \mathbf{B}) = \mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i),\ \mathbf{B} = B_1 \times \ldots \times B_n.$$

Soit S une fonction mesurable de n arguments. Alors $S = S(\mathbf{X}) = S(X_1, \ldots, X_n)$ est appelée **statistique**. Lorsqu'on effectue l'expérience on observe $s^{\mathrm{obs}} = S(\mathbf{x}^{\mathrm{obs}})$.

La loi Q_X est entièrement ou partiellement inconnue. Dans ce chapitre on étudiera l'estimation des paramètres inconnus. Ainsi si θ est un paramètre dont la loi Q_X dépende, on cherche une statistique fonction d'échantillon

$$\theta^* = \theta_n^*(\mathbf{X}).$$

Lorsqu'on observe l'échantillon $\mathbf{x}^{\mathrm{obs}} = (x_1^{\mathrm{obs}}, \ldots, x_n^{\mathrm{obs}})$, la valeur $\theta_n^*(\mathbf{x}^{\mathrm{obs}})$ est une **estimation** et devrait être assez proche de la vraie valeur du paramètre θ, lorsque n est assez grand. On verra plus loin quand cela est possible. La statistique $\theta_n^*(\mathbf{X})$ s'appelle **estimateur** du paramètre θ. Un estimateur est une règle de construction d'estimations. Nous appellerons estimateur θ_n^* les seules statistiques destinées à remplacer le paramètre inconnu θ.

Souvent, dans un problème d'estimation, on spécifie l'ensemble Θ des valeurs possibles du paramètre θ. C'est l'espace des paramètres. Aussi, dans des nombreux cas, on sait à l'avance que la loi Q_X de l'échantillon ne peut être arbitraire, mais appartient à une famille bien définie de lois $\mathcal{P}$.

Exemple. L'un des principaux paramètres caractérisant la qualité d'un système (machine, ampoule, ordinateur, etc) est la durée de service. Mais cette durée est en principe aléatoire et impossible à déterminer à l'avance. Il est toutefois raisonnable (si la fabrication est en quelque sorte homogène) de penser que les durées de service $X_1, X_2, \ldots$ sont des variables aléatoires indépendantes de même loi. Il est naturel d'identifier le paramètre durée de service au nombre $\theta = \mathrm{E}(X_i)$. On veut déterminer la valeur de θ. On observe n systèmes et on trouve $x_1^{\mathrm{obs}}, \ldots, x_n^{\mathrm{obs}}$. On sait que, lorsque $n \to \infty$,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{p.s.}} \theta.$$

Il est donc intuitif que le nombre $\bar{x}^{\mathrm{obs}} = \frac{1}{n}\sum_{i=1}^{n} x_i^{\mathrm{obs}}$ soit proche de θ pour n assez grand. Un exemple de famille de lois est $\mathcal{P} = \{\mathcal{E}(\lambda) : \lambda > 0\}$, $\theta = 1/\lambda \in \Theta =]0, \infty[$.

1.2 Familles paramétriques

1. Distribution gaussienne sur $\mathbb{R}$.
$\mathcal{N}(m, \sigma^2)$ de densité

$$g_{m,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

et

$$X \sim \mathcal{N}(m, \sigma^2) \Leftrightarrow X = \sigma G + m, \ \text{ avec } \ G \sim \mathcal{N}(0, 1)$$

On a

$$\mathrm{E}(e^{\lambda X}) = e^{\lambda m + \frac{\lambda^2\sigma^2}{2}},\ \mathrm{E}(e^{\lambda G}) = e^{\frac{\lambda^2}{2}},\ \mathrm{E}(G^{2k+1}) = 0,\ \mathrm{E}(G^{2k}) = \frac{2k!}{k!2^k}$$

et

$$\mathrm{E}(e^{i\lambda X}) = e^{i\lambda m - \frac{\lambda^2\sigma^2}{2}}.$$

2. Distribution gaussienne multidimensionnelle.
$\mathcal{N}_d(m, K)$, où K est une matrice $d \times d$ symétrique positive, c'est-à-dire telle que $\sum_{i,j=1}^{d} \theta_i K_{ij} \theta_j \geqslant 0$, et où $m \in \mathbb{R}^d$. On a

$$X \sim \mathcal{N}_d(m, K) \Leftrightarrow \mathrm{E}(e^{i<t,X>}) = e^{i<t,m> - \frac{1}{2}t^* K t}.$$

Si K est inversible, la densité est

$$g_{m,K}(x) := \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det K}} \exp\left(-\frac{1}{2}(x-m)^* K^{-1}(x-m)\right), \ \ x \in \mathbb{R}^d.$$

3. Distribution gamma.
$\gamma(p,\lambda)$, $p,\lambda>0$ de densité

$$\gamma_{p,\lambda}(x) := \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} \mathbb{1}_{x\geqslant 0},$$

où $\Gamma(p) := \int_0^\infty x^{p-1}e^{-x}dx$.
Si $X \sim \gamma(p,\lambda)$, alors

$$\mathrm{E}(e^{itX}) = \left(\frac{\lambda}{\lambda - it}\right)^p$$

et

$$\mathrm{E}(X) = \frac{p}{\lambda},\ \mathrm{E}(X^2) = \frac{p(p+1)}{\lambda^2},\ \mathrm{Var}(X) = \frac{p}{\lambda^2}.$$

4. Distribution de chi-deux.
$\chi^2(k)$, où $k \in \mathbb{N}^*$ de densité

$$f(x) := \frac{\left(\frac{1}{2}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} \mathbb{1}_{x>0}$$

Si $X = \xi_1^2 + \ldots + \xi_k^2$, avec ξ_j indépendantes de même loi $\mathcal{N}(0,1)$, alors $X \sim \chi^2(k) = \gamma(\frac{k}{2}, \frac{1}{2})$.
Si $X \sim \chi^2(k)$ et $Y \sim \chi^2(\ell)$ sont indépendantes alors $X+Y \sim \chi^2(k+\ell)$.
Si $X \sim \mathcal{N}_d(m,K)$ avec K inversible, alors $Q(X) := (X-m)^* K^{-1}(X-m) \sim \chi^2(d)$.
En effet,

$$\mathrm{E}\left(e^{\lambda Q(X)}\right) = (2\pi)^{-d/2}(\det K)^{-1/2} \int e^{\lambda Q(x) - \frac{1}{2}Q(x)} dx = (2\pi)^{-d/2}(\det K)^{-1/2} \int e^{-(-2\lambda+1)\frac{1}{2}Q(x)} dx.$$

Effectuant le changement de variable $\sqrt{1-2\lambda}(x_j - m_j) = y_j$ on trouve $\mathrm{E}\left(e^{\lambda Q(X)}\right) = (1-2\lambda)^{-d/2}$, puis faire $\lambda = it$ pour obtenir

$$\mathrm{E}\left(e^{itQ(X)}\right) = (1-2it)^{-d/2} = \left(\frac{1/2}{1/2 - it}\right)^{d/2}.$$

Notons aussi que si $X \sim \chi^2(k)$, alors $\mathrm{E}(X) = k$, $\mathrm{Var}(X) = 2k$, d'où

$$\frac{X-k}{\sqrt{2k}} \xrightarrow{\text{loi}} \mathcal{N}(0,1), \text{ pour } k \to \infty.$$

On peut aussi montrer que

$$\sqrt{2X} - \sqrt{2k-1} \xrightarrow{\text{loi}} \mathcal{N}(0,1), \text{ pour } k \to \infty.$$

5. Distribution exponentielle.
$\mathcal{E}(\lambda)$, $\lambda > 0$, de densité

$$f(x) := \lambda e^{-\lambda x} \mathbb{1}_{x \geqslant 0},$$

c'est-à-dire $\gamma(1, \lambda)$. On a $\mathrm{E}(X) = \frac{1}{\lambda}$ et $\mathrm{Var}(X) = \frac{1}{\lambda^2}$.

6. Distribution de Fisher.
Soient $k_1, k_2 \in \mathbb{N}^*$ et $\chi^2(k_1), \chi^2(k_2)$ indépendantes. Alors

$$\xi = \frac{\chi^2(k_1)/k_1}{\chi^2(k_2)/k_2}$$

suit une loi de Fisher $\mathcal{F}(k_1, k_2)$. La densité est, avec $p_j = \frac{k_j}{2}$, $j = 1, 2$,

$$f_\xi(t) = \frac{\Gamma(p_1 + p_2)}{\Gamma(p_1)\Gamma(p_2)} \frac{t^{p_1 - 1}}{[1 + (p_1/p_2)t]^{p_1 + p_2}} \mathbb{1}_{t > 0}.$$

On peut aussi voir que si $\xi_1, \ldots, \xi_{k_1}, \eta_1, \ldots, \eta_{k_2}$ sont des variables aléatoires gaussiennes centrées réduites indépendantes, alors

$$\xi \sim \frac{(\xi_1^2 + \ldots + \xi_{k_1}^2)/k_1}{(\eta_1^2 + \ldots + \eta_{k_2}^2)/k_2}.$$

On peut voir que

$$\mathrm{E}(\xi^\ell) = \frac{\Gamma(p_1 + \ell)\Gamma(p_2 - \ell)}{\Gamma(p_1)\Gamma(p_2)},$$

pour $\ell \in \left\{0, 1, \ldots, \left[\frac{k_2}{2}\right] - 1\right\}$. Par exemple $\mathrm{E}(\xi) = \frac{p_1}{p_2 - 1} = \frac{k_1}{k_2 - 2}$.
Enfin on peut voir que $1/F \sim \mathcal{F}(k_2, k_1)$.

7. Distribution de Student.
Soit $k \in \mathbb{N}^*$ et soient $\xi_0, \xi_1, \ldots, \xi_k$ des variables aléatoires gaussiennes centrées réduites indépendantes. Alors la variable aléatoire

$$T_k = \frac{\xi_0}{\sqrt{\frac{1}{k}(\xi_1^2 + \ldots + \xi_k^2)}}$$

suit la loi de Student à k degrés de liberté. On voit que T_k a une loi symétrique, et que

$$T_k^2 = \frac{k\xi_0^2}{\xi_1^2 + \ldots + \xi_k^2} \sim \mathcal{F}(1, k).$$

On en déduit

$$f_{T_k}(t) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \frac{1}{(1 + \frac{t^2}{k})^{\frac{k+1}{2}}}.$$

De plus
$$\mathrm{E}(T_k^{2\ell+1}) = 0,\ \mathrm{E}(T_k^{2\ell}) = k^\ell \frac{\Gamma(p_1+\ell)\Gamma(p_2-\ell)}{\Gamma(p_1)\Gamma(p_2)},$$
avec $p_1 = \frac{1}{2}$ et $p_2 = \frac{k}{2}$. Par exemple $\mathrm{E}(T_k^2) = k\frac{\Gamma(p_1+1)\Gamma(p_2-1)}{\Gamma(p_1)\Gamma(p_2)} = \frac{kp_1}{p_2-1} = \frac{k}{k-2}$.
Enfin, si on applique Stirling,
$$\frac{1}{\sqrt{k\pi}}\frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \to \frac{1}{\sqrt{2\pi}}, \text{ lorsque } k \to \infty$$
et aussi
$$(1+\frac{t^2}{k})^{-\frac{k+1}{2}} \to e^{-t^2/2}, \text{ lorsque } k \to \infty,$$
d'où
$$\lim_{k\to\infty} F_{T_k}(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}.$$

8. Distribution beta.
$\beta(p_1,p_2)$, avec $p_1,p_2 > 0$, de densité
$$f(x) = \frac{\Gamma(p_1+p_2)}{\Gamma(p_1)\Gamma(p_2)}x^{p_1-1}(1-x)^{p_2-1}\mathbb{1}_{[0,1]}(x).$$
On utilise souvent la fonction beta
$$B(p_1,p_2) = \frac{\Gamma(p_1)\Gamma(p_2)}{\Gamma(p_1+p_2)}$$
Soient $\eta_j \sim \gamma(p_j,\lambda)$, $j = 1,2$. Alors $\beta = \frac{\eta_1}{\eta_1+\eta_2} \sim \beta(p_1,p_2)$. En effet
$$\mathrm{P}(\beta \leqslant t) = \mathrm{P}(\frac{\eta_1}{\eta_1+\eta_2} \leqslant t) = \mathrm{P}(\frac{\eta_1/\eta_2}{1+\eta_1/\eta_2} \leqslant t) = \mathrm{P}(\frac{\xi}{1+\xi} \leqslant t) = \mathrm{P}(\xi \leqslant \frac{t}{1-t})$$
avec $\xi \sim \mathcal{F}(p_1,p_2)$. On trouve
$$f_\beta(t) = \frac{\left(\frac{t}{1-t}\right)^{p_1-1}}{\left(1+\frac{t}{1-t}\right)^{p_1+p_2}}\frac{\Gamma(p_1+p_2)}{\Gamma(p_1)\Gamma(p_2)}\frac{1-t+t}{(1-t)^2}$$
$$= \left(\frac{t}{1-t}\right)^{p_1-1}(1-t)^{p_1+p_2-2}\frac{\Gamma(p_1+p_2)}{\Gamma(p_1)\Gamma(p_2)} = t^{p_1-1}(1-t)^{p_2-1}\frac{\Gamma(p_1+p_2)}{\Gamma(p_1)\Gamma(p_2)}.$$
De plus
$$\mathrm{E}(\beta^\ell) = \frac{\Gamma(p_1+p_2)\Gamma(p_1+\ell)}{\Gamma(p_1)\Gamma(p_1+p_2+\ell)}.$$
Par exemple $\mathrm{E}(\beta) = \frac{p_1}{p_1+p_2}$ et $\mathrm{E}(\beta^2) = \frac{p_1(p_1+1)}{(p_1+p_2)(p_1+p_2+1)}$.

9. Distribution uniforme.
$\mathcal{U}_{[a,b]}$, avec $a < b$, de densité

$$f(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x).$$

On a $\mathrm{E}(X) = \frac{a+b}{2}$ et $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$.
Pour $a = 0, b = 1$ il s'agit de la distribution $\beta(1,1)$.

10. Distribution de Cauchy.
$\mathcal{C}(\alpha, \sigma^2)$, avec $\alpha \in \mathbb{R}$ et $\sigma^2 > 0$, de densité

$$f(x) = \frac{1}{\pi\sigma}\frac{1}{1+\left(\frac{x-\alpha}{\sigma}\right)^2}.$$

Si $X_0 \sim \mathcal{C}(0,1)$ alors $X = \alpha + \sigma X_0 \sim \mathcal{C}(\alpha, \sigma^2)$.
On a

$$\mathrm{E}(e^{itX_0}) = e^{-|t|},\ \mathrm{E}(e^{itX}) = e^{it\alpha - \sigma|t|}.$$

Enfin, si $X \sim \mathcal{C}(\alpha, \sigma^2)$ et $Y \sim \mathcal{C}(\beta, \tau^2)$ sont indépendantes alors $X+Y \sim \mathcal{C}(\alpha+\beta, (\sigma+\tau)^2)$.

11. Distribution log-normale.
η est log-normale si $\log \eta \sim \mathcal{N}(m, \sigma^2)$, ou $\eta = e^{\xi}$ avec $\xi \sim \mathcal{N}(m, \sigma^2)$; $\eta > 0$.
On écrit, pour $t > 0$, $\mathrm{P}(\eta \leqslant t) = \mathrm{P}(e^{\xi} \leqslant t) = \mathrm{P}(\xi \leqslant \log t)$, d'où

$$f_\eta(t) = g_{m,\sigma^2}(\log t)\frac{1}{t}\mathbb{1}_{t>0}.$$

On a $\mathrm{E}(\eta) = e^{m+\frac{\sigma^2}{2}}$, $\mathrm{E}(\eta^2) = e^{2m+2\sigma^2}$ et $\mathrm{Var}(\eta) = e^{2m+\sigma^2}(e^{\sigma^2} - 1)$.

12. Loi binomiale.
$\mathcal{B}(n,p)$. On a $\mathrm{P}(X = k) = C_n^k p^k (1-p)^{n-k}$, $k = 0, \ldots, n$, $p \in]0,1[$. Aussi $\mathrm{E}(X) = np$, $\mathrm{Var}(X) = np(1-p)$, $\mathrm{E}(e^{itX}) = (1-p+pe^{it})^n$.

13. Distribution de Poisson.
$\mathcal{P}(\lambda)$. On a $\mathrm{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$, $k \in \mathbb{N}$, où $\lambda > 0$.
Aussi $\mathrm{E}(X) = \mathrm{Var}(X) = \lambda$, $\mathrm{E}(e^{itX}) = \exp(\lambda(e^{it} - 1))$.

14. Distribution géométrique.
$\mathcal{G}(p)$. On a $\mathrm{P}(X = k) = p(1-p)^{k-1}$, $k \geqslant 1$, $p \in]0,1[$. Aussi $\mathrm{E}(X) = \frac{1}{p}$, $\mathrm{Var}(X) = \frac{1-p}{p^2}$, $\mathrm{E}(e^{itX}) = \frac{pe^{it}}{1-(1-t)e^{it}}$.

15. Distribution uniforme.
$\mathcal{U}(1, \ldots, r)$. On a $\mathrm{P}(X = x_j) = \frac{1}{r}$, $j = 1, \ldots, r$. $\mathrm{E}(X) = \frac{r+1}{2}$, $\mathrm{Var}(X) = \frac{(r+1)(2r+1)}{6}$ et $\varphi_X(t) = \frac{1}{r}e^{it}\frac{1-e^{irt}}{1-e^{it}}$.

16. Distribution multinomiale.
$\mathcal{M}(n, p_1, \ldots, p_n)$. Soient $0 < p_j < 1$, $j = 1, \ldots, r$ tels que $p_1 + \ldots + p_r = 1$. Pour $k = (k_1, \ldots, k_r)$, tel que $k_1 + \ldots + k_r = n$ on définit la loi de $(X_1, \ldots, X_r)$ par

$$\mathrm{P}((X_1, \ldots, X_r) = (k_1, \ldots, k_r)) = \frac{n!}{k_1! \ldots k_r!} p_1^{k_1} \ldots p_r^{k_r}.$$

On a

$$\mathrm{E}(X_1, \ldots, X_r) = (np_1, \ldots, np_r),$$
$$\mathrm{Var}(X_j) = np_j(1 - p_j),\ \mathrm{Cov}(X_j, X_k) = -np_j p_k$$

et

$$\varphi_X(t_1, \ldots, t_r) = \left(\sum_{j=1}^{r} p_j e^{it_j p_j} \right)^n.$$

1.3 Distribution empirique

Définition 1.1 *1) Soit $\{X_n : n \geqslant 1\}$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$ et X une variable aléatoire à valeurs dans $\mathbb{R}^d$. On dit que X_n* **converge en loi** *vers X lorsque $n \to \infty$, si pour toute fonction $g : \mathbb{R}^d \to \mathbb{R}$ continue et bornée on a*

$$\lim_{n \to \infty} \mathrm{E}\left[g(X_n)\right] = \mathrm{E}\left[g(X)\right].$$

2) Soit $\{\mu_n : n \geqslant 1\}$ une suite de probabilités sur $\mathcal{B}(\mathbb{R}^d)$ et μ une probabilité sur $\mathbb{R}^d$. On dit que μ_n **converge étroitement** *vers μ, si pour toute fonction $g : \mathbb{R}^d \to \mathbb{R}$ continue et bornée on a*

$$\lim_{n \to \infty} \int g(x) \mu_n(dx) = \int g(x) \mu(dx).$$

3) Si X_n est de loi Q_n pour tout n et si X est de loi Q, on voit que la convergence en loi de la suite X_n vers X est **équivalente** *à la convergence étroite de la suite Q_n vers Q.*

4) Soit $\{X_n : n \geqslant 1\}$ une suite de variables aléatoires de même loi Q et indépendantes. Soit $\hat{Q}_n$ la mesure aléatoire suivante dite **distribution empirique** *de l'échantillon :*

$$\hat{Q}_n(\omega) = \frac{1}{n}(\delta_{X_1(\omega)} + \ldots + \delta_{X_n(\omega)}),\ \omega \in \Omega,$$

où $\delta_{X_i(\omega)}$ est la masse de Dirac en $X_i(\omega)$.

Le fait remarquable est que cette suite de mesures aléatoires se rapproche indéfiniment de la loi Q de la variable aléatoire observée.

Théorème 1.1 *Soit $\{X_n : n \geqslant 1\}$ une suite de variables aléatoires indépendantes et de même loi Q. Lorsque $n \to \infty$,*

$$\hat{Q}_n(B) \xrightarrow{p.s.} Q(B),\ \forall B \in \mathcal{B}(\mathbb{R}^d).$$

Preuve. Les variables aléatoires $\mathbb{1}_B(X_i) \sim \mathcal{B}(1, Q(B))$ sont indépendantes de même loi. Le résultat s'obtient par la loi forte des grands nombres. $\square$

Remarque : Il s'agit d'une convergence en chaque "point" B. On peut faire mieux. Dans le résultat qui suit supposons que la dimension est $d = 1$: on note $\mathcal{J} = \{]a, b]\}$, $-\infty \leqslant a < b \leqslant \infty$. Alors, lorsque $n \to \infty$,

$$\sup_{B \in \mathcal{J}} |\hat{Q}_n(B) - Q(B)| \xrightarrow{\text{p.s.}} 0.$$

$\square$

Au lieu de prouver ce résultat on va prouver un résultat équivalent sur la **fonction de répartition empirique** : pour $d = 1$, on note la variable aléatoire

$$\hat{F}_n(t) := \hat{Q}_n(]-\infty, t]) = \frac{1}{n}\text{card}\{i : 1 \leqslant i \leqslant n, X_i \leqslant t\} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{]-\infty,t]}(X_i),\ t \in \mathbb{R}.$$

On notera la fonction de répartition de la variable X :

$$F(t) = Q(]-\infty, t]),\ t \in \mathbb{R},$$

mais aussi

$$\hat{F}_n(t-) := \hat{Q}_n(]-\infty, t[),\ F(t-) = Q(]-\infty, t[),\ t \in \mathbb{R}.$$

Théorème 1.2 *(Glivenko-Cantelli)*
Soit $\{X_n : n \geqslant 1\}$ une suite de variables aléatoires réelles indépendantes et de même loi Q. Alors, lorsque $n \to \infty$,

$$\sup_{t \in \mathbb{R}} |\hat{F}_n(t) - F(t)| \xrightarrow{p.s.} 0.$$

Preuve. Par le Théorème 1.1 appliqué au $B =]-\infty, t]$ ensuite au $C =]-\infty, t[$, on trouve Ω'_t, Ω''_t événements de probabilité 1, tels que si $n \to \infty$ on a

$$\hat{F}_n(t, \omega) \to F(t),\ \forall \omega \in \Omega'_t$$

et

$$\hat{F}_n(t-, \omega) \to F(t-)\ \forall \omega \in \Omega''_t.$$

Notons $G(u) = \inf\{t : F(t) \geqslant u\}$, pour $0 < u < 1$, l'inverse généralisée à droite de F. On a

$$F(G(u)-) \leqslant u \leqslant F(G(u)).$$

Pour $m \in \mathbb{N}^*$ et $1 \leqslant k \leqslant m$, on pose $t_k^m := G(\frac{k}{m})$. Si $t_{k-1}^m \leqslant t < t_k^m$, alors, par monotonie,

$$F(t_{k-1}^m) \leqslant F(t) \leqslant F(t_k^m-) \text{ et } \hat{F}_n(t_{k-1}^m) \leqslant \hat{F}_n(t) \leqslant \hat{F}_n(t_k^m-).$$

Pour des tels t,

$$\hat{F}_n(t_{k-1}^m, \omega) - F(t_k^m-) \leqslant \hat{F}_n(t, \omega) - F(t) \leqslant \hat{F}_n(t_k^m-, \omega) - F(t_{k-1}^m).$$

Mais

$$F(t_k^m-) - F(t_{k-1}^m) \leqslant \frac{1}{m},$$

ainsi que

$$F(t_1^m-) \leqslant \frac{1}{m},\ F(t_m^m) \geqslant 1 - \frac{1}{m}.$$

Donc

$$\hat{F}_n(t_{k-1}^m, \omega) - F(t_{k-1}^m) - \frac{1}{m} \leqslant \hat{F}_n(t, \omega) - F(t) \leqslant \hat{F}_n(t_k^m-, \omega) - F(t_k^m-) + \frac{1}{m},$$

ou encore

$$\sup_{t \in [t_{k-1}^m, t_k^m[} |\hat{F}_n(t, \omega) - F(t)| \leqslant \max\{|\hat{F}_n(t_{k-1}^m, \omega) - F(t_{k-1}^m)|, |\hat{F}_n(t_k^m-, \omega) - F(t_k^m-)|\}.$$

On note $D_n(\omega) = \sup_{t \in \mathbb{R}} |\hat{F}_n(t, \omega) - F(t)|$ et

$$D_n^m(\omega) = \max_{k=1,\dots,m} \max\{|\hat{F}_n(t_k^m, \omega) - F(t_k^m)|, |\hat{F}_n(t_k^m-, \omega) - F(t_k^m-)|\}.$$

Alors

$$\max_{k=2,\dots,m} \sup_{t \in [t_{k-1}^m, t_k^m[} |\hat{F}_n(t, \omega) - F(t)| \leqslant D_n^m(\omega) + \frac{1}{m},$$

Donc

$$\sup_{t \in [t_1^m, t_m^m[} |\hat{F}_n(t, \omega) - F(t)| \leqslant D_n^m(\omega) + \frac{1}{m}$$

Par des arguments similaires pour $t < t_1^m$ et $t \geqslant t_m^m$ on trouve

$$D_n(\omega) \leqslant D_n^m(\omega) + \frac{1}{m}.$$

Lorsque $\omega \in \Omega_m := \bigcap_k \Omega'_{t_k^m} \cap \bigcap_k \Omega''_{t_k^m}$, on a $\lim_{n\to\infty} D_n^m(\omega) = 0$, d'où on déduit que $\limsup_{n\to\infty} D_n(\omega) \leqslant \frac{1}{m}$. De plus il est facile de voir que Ω_m est de probabilité 1. Enfin, pour $\omega \in \bar{\Omega} = \bigcap_m \Omega_m$ on déduit $\lim_{n\to\infty} D_n(\omega) = 0$ avec $\bar{\Omega}$ de probabilité 1, d'où la conclusion. $\square$

Remarque : Le théorème de Glivenko-Cantelli est un cas particulier du résultat de la remarque précédente, car $]-\infty, t] \in \mathcal{J}$; réciproquement, si on prend $B =]a, b]$,

$$|\hat{\mathrm{P}}_n(B) - \mathrm{P}(B)| \leqslant |\hat{F}_n(b) - F(b)| + |\hat{F}_n(a) - F(a)|$$

si bien que

$$\sup_{B \in \mathcal{J}} |\hat{\mathrm{P}}_n(B) - \mathrm{P}(B)| \leqslant \sup_{a,b} [|\hat{F}_n(b) - F(b)| + |\hat{F}_n(a) - F(a)|] \xrightarrow{\text{p.s.}} 0.$$

□

On revient au cas de la dimension d quelconque. La meilleure convergence est contenue dans le résultat suivant :

Théorème 1.3 *Soit $\{X_n : n \geqslant 1\}$ une suite de variables aléatoires indépendantes et de même loi Q. Presque sûrement, lorsque $n \to \infty$,*

$$\hat{Q}_n \xrightarrow{e} Q.$$

Preuve. Pour toute fonction continue et bornée :

$$\int g(x)\hat{Q}_n(dx) = \frac{1}{n}(g(X_1) + \ldots + g(X_n))$$

et le membre de droite converge presque sûrement vers $\mathrm{E}[g(X)]$ d'après la loi forte des grands nombres. Comme

$$\mathrm{E}[g(X)] = \int g(x)Q(dx),$$

le résultat est prouvé. □

Remarque : Le même résultat reste vrai pour la convergence de distributions empiriques définie à l'aide de toute fonction g telle que $\int |g(x)|Q(dx) < \infty$. □

Théorème 1.4 *Soit $\{X_n : n \geqslant 1\}$ une suite de variables aléatoires indépendantes et de même loi Q. Soit $g : \mathbb{R}^d \to \mathbb{R}$ borélienne telle que $\int |g(x)|Q(dx) < \infty$ (c'est-à-dire $g \in \mathrm{L}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), Q)$) et soit $h : \mathbb{R} \to \mathbb{R}$ continue au point $m_g = \int g(x)Q(dx) = \mathrm{E}[g(X)]$. Définissons*

$$\Phi_{g,h}(Q) := h\left(\int g(x)Q(dx)\right) = h\left(\mathrm{E}[g(X)]\right) = h(m_g).$$

et on note

$$\Phi_{g,h}(\hat{Q}_n) = h\left(\int g(x)\hat{Q}_n(dx)\right) = h\left(\frac{1}{n}(g(X_1) + \ldots + g(X_n))\right).$$

Lorsque $n \to \infty$,

$$\Phi_{g,h}(\hat{Q}_n) \xrightarrow{p.s.} \Phi_{g,h}(Q).$$

Preuve. D'après la loi des grands nombres, presque sûrement

$$\lim_{n\to\infty} \frac{1}{n}(g(X_1) + \ldots + g(X_n)) = \int g(x)Q(dx).$$

Il existe Ω_0, de probabilité 1 tel que

$$\forall \omega \in \Omega_0,\ \forall \delta > 0,\ \exists n_0(\delta, \omega) \text{ t.q. } \forall n \geqslant n_0 : \left| \frac{1}{n} \sum_{j=1}^{n} g(X_j(\omega)) - \int g(x)Q(dx) \right| < \delta.$$

Soit $\varepsilon > 0$; il existe $\delta > 0$ tel que dès que $|x - m_g| < \delta$, on a $|h(x) - h(m_g)| < \varepsilon$. D'où, pour n assez grand ($n \geqslant n_0$ avec n_0 dépendant de $\omega \in \Omega_0$),

$$\left| h\left(\frac{1}{n} \sum_{j=1}^{n} g(X_j(\omega)) \right) - h\left(\int g(x)Q(dx) \right) \right| < \varepsilon.$$

$\square$

Définition 1.2 *Soit* $\{X_n : n \geqslant 1\}$ *une suite de variables aléatoires réelles indépendantes et de même loi* Q. *Les moments et les moments centrés de la distribution empirique sont nommés* **moments et moments centrés empiriques**. *Ainsi,*

$$\hat{m}_k = \int x^k \hat{Q}_n(dx) = \int x^k \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}(dx) = \frac{1}{n} \sum_{i=1}^{n} X_i^k.$$

Le plus important est **la moyenne empirique** $\bar{X} = \hat{m}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i$. *Ensuite*

$$\hat{m}_k^0 = \int (x - \bar{X})^k \hat{Q}_n(dx) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^k.$$

Le plus important est **la variance empirique** $S^2 = \hat{m}_2^0 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$.

Corollaire 1.1 *Soit* $\{X_n : n \geqslant 1\}$ *une suite de variables aléatoires réelles indépendantes et de même loi* Q. *Supposons que* X *est une variable aléatoire de loi* Q *qui admet des moments d'ordre deux. Lorsque* $n \to \infty$,

$$\bar{X} \xrightarrow{p.s.} \mathrm{E}(X),\ S^2 \xrightarrow{p.s.} \mathrm{Var}(X),$$

où

$$\mathrm{E}(X) = \int xQ(dx),\ \mathrm{Var}(X) = \int x^2 Q(dx) - \left(\int xQ(dx) \right)^2.$$

On peut préciser aussi des convergences en loi de ce type de fonctionnelles :

Théorème 1.5 *Soit $\{X_n : n \geqslant 1\}$ une suite de variables aléatoires indépendantes et de même loi Q. On suppose que $\int g^2(x)Q(dx) < \infty$ (c'est-à-dire que $g \in \mathrm{L}^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), Q)$) et que $h : \mathbb{R} \to \mathbb{R}$ est de classe C^2 avec $h'(m) \neq 0$ et sa dérivée seconde bornée. On note $\sigma_g^2 = \int g^2(x)Q(dx) - \left(\int g(x)Q(dx)\right)^2 = \mathrm{Var}[g(X)]$. Alors, lorsque $n \to \infty$,*

$$\sqrt{n}\left(\Phi_{g,h}(\hat{Q}_n) - \Phi_{g,h}(Q)\right) \xrightarrow{loi} \mathcal{N}(0, \sigma_g^2(h'(m_g))^2).$$

La preuve de ce résultat s'obtient à l'aide des trois lemmes suivants.

Lemme 1.1 *Si une suite $\xi_n \xrightarrow{loi} \xi$ et si la suite $\alpha_n \xrightarrow{proba} 0$, alors $\alpha_n\xi_n \xrightarrow{loi} 0$.*

Preuve. Soient $\varepsilon, \eta > 0$ et t assez grand tel que $\mathrm{P}(|\xi| \geqslant t) < \eta$. Soit $n_0 \in \mathbb{N}$ tel que si $n \geqslant n_0$,

$$\mathrm{P}(|\alpha_n| > \delta) < \eta \text{ et } |\mathrm{P}(\xi_n \geqslant s) - \mathrm{P}(\xi \geqslant s)| < \eta, \text{ pour tout } s.$$

Alors, si $n \geqslant n_0$, alors $\mathrm{P}(|\alpha_n\xi_n| \geqslant \varepsilon) < 3\eta$. On en déduit que $\alpha_n\xi_n$ converge en probabilité vers 0, donc en loi. □

Lemme 1.2 *(***théorème central limite***)*

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} g(X_j) - \int g(x)Q(dx)\right) = \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} g(X_j) - m_g\right) \xrightarrow{loi} \mathcal{N}(0, \sigma_g^2)$$

Lemme 1.3 *Si une suite $\xi_n \xrightarrow{loi} \xi$ et si ϕ est continue, alors $\phi(\xi_n) \xrightarrow{loi} \phi(\xi)$.*

Preuve. Soit ψ fonction continue bornée quelconque et on montre que $\mathrm{E}[\psi(\phi(\xi_n))] \to \mathrm{E}[\psi(\phi(\xi))]$. Mais cela est une conséquence de la convergence en loi de ξ_n vers ξ puisque la fonction $\psi \circ \phi$ est continue bornée. □

Preuve du Théorème 1.5 On continue de noter $m_g = \int g(x)Q(dx) = \mathrm{E}[g(X)]$. écrivons la formule de Taylor

$$\sqrt{n}\left(h\left(\frac{1}{n}\sum_{j=1}^{n} g(X_j)\right) - h(m_g)\right) = \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} g(X_j) - m_g\right)h'(m_g)$$

$$+\frac{\sqrt{n}}{2}\left(\frac{1}{n}\sum_{j=1}^{n} g(X_j) - m_g\right)^2 h''(\bar{m}),$$

où $\bar{m}(\omega)$ est un point situé entre m_g et $\frac{1}{n}\sum_{j=1}^{n} g(X_j(\omega))$. écrivant le second terme de cette expression comme

$$\frac{1}{2\sqrt{n}}\left(\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} g(X_j) - m_g\right)\right)^2 h''(\bar{m}),$$

le théorème résulte des lemmes précédents. En effet, comme la dérivée seconde de h est bornée, on peut majorer cette dernière expression par une quantité qui converge en loi vers zéro en appliquant Lemmes 1.1 et 1.3 avec

$$\alpha_n = \frac{1}{2\sqrt{n}},\ \xi_n = \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} g(X_j) - m_g\right),\ \xi \sim \mathcal{N}(0, \sigma_g^2) \text{ et } \phi(x) = x^2.$$

Le premier terme du membre droit dans le développement de Taylor converge en loi vers la loi gaussienne de l'énoncé. □

Remarque : Il suffit en fait pour ce théorème de supposer que h est dérivable avec dérivée continue en m.

1.4 Méthodes d'estimation ponctuelle

On indiquera essentiellement deux méthodes de construction d'estimateurs. On considère un n-échantillon $X_1, X_2, \ldots, X_n$ de loi appartenant à une famille paramétrique $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\}$. On notera la vraie valeur du paramètre (inconnue) θ_0 et le but est de construire des estimateurs de θ_0, fonctions mesurables d'échantillon : $\theta_n^* = \theta^*(\mathbf{X})$. On essaie aussi d'assurer des bonnes propriétés asymptotiques de ces estimateurs.

Définition 1.3 *Une suite d'estimateurs* $\{\theta_n^* : n \geqslant 1\}$ *est dite* **convergente (ou consistante)** *si, lorsque* $n \to \infty$*, sous* P_θ,

$$\theta_n^* \xrightarrow{prob} \theta,\ \forall \theta \in \Theta.$$

Elle est dite **fortement convergente** *si, lorsque la précédente convergence a lieu presque sûrement. Elle est dite* **asymptotiquement normale** *si lorsque* $n \to \infty$*, sous* P_θ,

$$\sqrt{n}(\theta_n^* - \theta) \xrightarrow{loi} \mathcal{N}(0, \sigma^2(\theta)),\ \forall \theta \in \Theta,$$

où $\sigma^2(\theta)$ *est une quantité à préciser.*

1.4.1 Méthode des moments

Il s'agit de décrire une méthode d'estimation basée sur la substitution d'une distribution empirique. Si X est de loi P_θ et si on sait qu'il existe une fonctionnelle Φ telle que $\theta = \Phi(\mathrm{P}_\theta)$, alors la **méthode de substitution** nous dit de prendre pour estimateur $\theta_n^* = \Phi(\hat{Q}_n)$. En général la fonctionnelle peut ne pas être définie sur l'ensemble de distributions empiriques. Dans le cas d'un modèle paramétrique on utilisera un cas particulier de cette méthode : **méthode des moments**.

Définition 1.4 *Supposons qu'il existe deux fonctions $g : \mathbb{R}^d \to \mathbb{R}$ et $q : \Theta \to \mathbb{R}$ telles que les quantités ci-dessus existent et sont finies :*

$$q(\theta) = \int g(x)\mathrm{P}_\theta(dx).$$

Si on suppose que la fonction q est bijective (par exemple, dans le cas $k = 1$, si q est monotone continue), alors,

$$\theta = q^{-1}\left(\int g(x)\mathrm{P}_\theta(dx)\right) = \Phi_{g,q^{-1}}(\mathrm{P}_\theta).$$

L'estimateur par la méthode des moments *(s'il est bien défini) est :*

$$\theta_n^* = \theta_n^*(\mathbf{X}) = \Phi_{g,q^{-1}}(\hat{Q}_n) = q^{-1}\left(\int g(x)d\hat{Q}_n(x)\right) = q^{-1}\left(\frac{1}{n}\sum_{j=1}^n g(X_j)\right).$$

Le nom de cette méthode vient du fait qu'elle consiste è a utiliser les moments empiriques. En effet l'estimateur n'est d'autre que la solution de l'équation

$$q(\theta) = \frac{1}{n}\sum_{j=1}^n g(X_j)$$

et souvent $g(x) = x$ ou $g(x) = x^r$.

Les estimateurs obtenus par la méthode des moments ont des bonnes propriétés asymptotiques :

Théorème 1.6 *1) (***convergence forte***) Supposons que $g : \mathbb{R}^d \to \mathbb{R}$ est telle que $\int g(x)\mathrm{P}_\theta(dx) < \infty$, pour tout $\theta \in \Theta$, et que $q^{-1} : \mathbb{R} \to \Theta$ est continue. Alors, l'estimateur par méthode de moments, lorsque $n \to \infty$,*

$$\theta_n^* = \theta_n^*(\mathbf{X}) \xrightarrow{p.s.} \theta, \forall \theta \in \Theta.$$

*2) (dimension 1)(***normalité asymptotique***) Supposons que $q : \Theta \subset \mathbb{R} \to \mathbb{R}$ est de classe C^1 avec $q' \neq 0$, et que $\int g^2(x)\mathrm{P}_\theta(dx) < \infty$, pour tout $\theta \in \Theta$. Supposons que $\mathrm{Var}_\theta[g(X)] = \int g^2(x)\mathrm{P}_\theta(dx) - \left(\int g(x)\mathrm{P}_\theta(dx)\right)^2 > 0$, pour tout $\theta \in \Theta$. Alors l'estimateur par la méthode des moments $\theta_n^* = \theta_n^*(\mathbf{X})$ est asymptotiquement normal :*

$$\sqrt{n}(\theta_n^* - \theta) \xrightarrow{loi} \mathcal{N}(0, \mathrm{Var}_\theta[g(X)]\frac{1}{q'(\theta)^2}), \forall \theta \in \Theta.$$

C'est une application directe des théorèmes 1.4 et 1.5.

Exemples :

i) $X \sim \mathcal{E}(\lambda)$. On a $\int x\mathrm{P}_\lambda(dx) = \frac{1}{\lambda}$ et $\int x^2\mathrm{P}_\lambda(dx) = \frac{2}{\lambda^2}$.

a) Soit $g(x) = x$, $\varphi(x) = \frac{1}{x}$, $\varphi^{-1}(y) = \frac{1}{y}$. Ici $\mathrm{Var}_\lambda(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$. On a

$$\lambda_n^* = \frac{n}{X_1 + \ldots + X_n},$$

d'où

$$\sqrt{n}\left(\frac{n}{X_1 + \ldots + X_n} - \lambda\right) \xrightarrow{\text{loi}} \mathcal{N}(0, \frac{1}{\lambda^2}\lambda^4) = \mathcal{N}(0, \lambda^2).$$

b) Soit $g(x) = x^2$, $\varphi^{-1}(y) = \sqrt{\frac{2}{y}}$. On a

$$\lambda_n^{**} = \left(\frac{2n}{X_1^2 + \ldots + X_n^2}\right)^{\frac{1}{2}},$$

d'où

$$\sqrt{n}(\lambda_n^{**} - \lambda) \xrightarrow{\text{loi}} \mathcal{N}(0, \mathrm{Var}_\lambda(X_1^2)\frac{\lambda^6}{16}) = \mathcal{N}(0, \frac{20}{\lambda^4} \cdot \frac{\lambda^6}{16}) = \mathcal{N}(0, \frac{5}{4}\lambda^2).$$

ii) $X \sim \mathcal{N}(m, \sigma^2)$. On a $\int x \mathrm{P}_{m,\sigma^2}(dx) = m$, $\int x^2 \mathrm{P}_{m,\sigma^2}(dx) = \sigma^2 + m^2$.

a) Soit σ^2 connu, par exemple $\sigma^2 = 1$. Soit $g(x) = x$, $\varphi(x) = x$. On a

$$m^* = \frac{1}{n}\sum_{j=1}^n X_j = \bar{X},$$

d'où

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^n X_j - m\right) \xrightarrow{\text{loi}} \mathcal{N}(0, 1).$$

En fait on a ici $\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^n X_j - m\right) \sim \mathcal{N}(0, 1)$.

écrire les résultats pour σ^2 quelconque (connu).

b) Soit m connu, par exemple $m = 0$. Soit $g(x) = x^2$, $\varphi(x) = x$. On a

$$(\sigma^2)^* = \frac{1}{n}\sum_{j=1}^n X_j^2,$$

d'où

$$\sqrt{n}\left((\sigma^2)^* - \sigma^2\right) \xrightarrow{\text{loi}} \mathcal{N}(0, 2).$$

écrire les résultats pour m quelconque (connu).

c) Les deux paramètres sont inconnus. On a

$$m^* = \frac{1}{n}\sum_{j=1}^n X_j = \bar{X},\ (\sigma^2 + m^2)^* = \frac{1}{n}\sum_{j=1}^n X_j^2,$$

et

$$(\sigma^2)^* = \frac{1}{n}\sum_{j=1}^n X_j^2 - \left(\frac{1}{n}\sum_{j=1}^n X_j\right)^2 = S^2.$$

1.4.2 Méthode du maximum de vraisemblance

Soit $\mathcal{P} = (\mathrm{P}_\theta : \theta \in \Theta)$ une famille paramétrique de distributions. On suppose :

$$(\mathrm{A}_0) \qquad \mathrm{P}_{\theta_1} \neq \mathrm{P}_{\theta_2} \text{ si } \theta_1 \neq \theta_2;$$

on dit que le modèle est **identifiable** ;

$$(\mathrm{A}_\mu) \ \exists \mu \text{ une mesure } \sigma\text{-finie sur } (E, \mathcal{B}(E)) \text{ t. q. } \forall \theta, \mathrm{P}_\theta \text{ est absolument continue}$$

$$\text{par rapport à } \mu, \text{ c'est-à-dire } \mathrm{P}_\theta \text{ admet une densité } f_\theta : \mathrm{P}_\theta(dx) = f_\theta(x) \cdot \mu(dx).$$

On dit que μ **domine** toutes les probabilités P_θ. En d'autres termes, pour tout borélien $B \in \mathcal{B}(E)$

$$\mathrm{P}_\theta(B) = \int_B f_\theta(x)\mu(dx), \ \forall \theta \in \Theta.$$

La fonction $f : \Theta \times E \to [0, \infty[$ est dite s'appelle **vraisemblance**. La fonction densité d'un n-échantillon $\mathbf{X} = (X_1, \dots, X_n)$ est

$$f_\theta(\mathbf{x}) = \prod_{j=1}^n f_\theta(x_j)$$

et continuera l'appeler **vraisemblance**.

Tous les exemples vus précédemment satisfont à cette hypothèse, avec μ la mesure de Lebesgue ou la mesure de comptage sur un ensemble dénombrable μ telle que $\mu(B) = k$ si B contient exactement k points de coordonnées entières). Cette hypothèse permet de traiter d'une façon unifiée les lois continues et les lois discrètes. De plus la dimension de l'espace E devient inessentielle.

Définition 1.5 *On appelle* **estimateur par la méthode du maximum de vraisemblance** *l'estimateur $\hat{\theta}_n^* = \hat{\theta}^*(\mathbf{X})$ (s'il existe) tel que*

$$\hat{\theta}^* = \hat{\theta}^*(\mathbf{x}) = \mathrm{argsup} f_\theta(\mathbf{x}),$$

c'est-à-dire

$$f_{\hat{\theta}^*}(\mathbf{x}) = \sup_{\theta \in \Theta} f_\theta(\mathbf{x}) = \sup_{\theta \in \Theta} \prod_{j=1}^n f_\theta(x_j).$$

ou, équivalent

$$\ln f_{\hat{\theta}^*}(\mathbf{x}) = \sup_{\theta \in \Theta} \ln f_\theta(\mathbf{x}) = \sup_{\theta \in \Theta} \sum_{j=1}^n \ln f_\theta(x_j).$$

Remarques :

1. Pour donner une interprétation intuitive de l'estimateur de maximum de vraisemblance supposons que la loi P_θ est discrète. Alors, si l'échantillon est de loi P_θ,

$$f_\theta(\mathbf{x}^{\text{obs}}) = P_\theta(X_1 = x_1^{\text{obs}}, \ldots, X_n = x_n^{\text{obs}}),$$

c'est-à-dire que $f_\theta(\mathbf{x}^{\text{obs}})$ est la probabilité d'observer effectivement $\mathbf{x}^{\text{obs}}$. Ainsi on estime le paramètre par la valeur de θ qui maximise la probabilité d'observer certaines valeurs données. Une interprétation similaire s'obtient pour le cas à densité par rapport à la mesure de Lebesgue.

2. L'estimateur de maximum de vraisemblance n'est pas unique car une fonction peut atteindre son maximum en plusieurs points. De plus l'estimateur de maximum de vraisemblance est totalement indépendant du choix de μ, puisque la substitution de μ par une mesure équivalente μ_0 se traduit par la multiplication de la fonction vraisemblance $f_\theta(\mathbf{x})$ par un facteur $\frac{\mu_0^n(dx_1 \ldots dx_n)}{\mu^n(dx_1 \ldots dx_n)}$ indépendant de θ.

3. Dans plusieurs cas importants il existe un unique estimateur de maximum de vraisemblance et il peut être obtenu par dérivation.

Exemples :

i) Soient $X_1, \ldots, X_n$ sont n variables aléatoires indépendantes de densité $\gamma_\lambda = \gamma_{p,\lambda}$ avec le paramètre p connu, on a pour $x_j > 0$, $j = 1, \ldots, n$,

$$\ln f_\lambda(\mathbf{x}) = \ln \prod_{j=1}^{n} \gamma_\lambda(x_j) = np \ln \lambda - n \ln \Gamma(p) + (p-1) \sum_{j=1}^{n} \ln x_j - \alpha \sum_{j=1}^{n} x_j.$$

On voit que $\frac{d}{d\lambda} \ln f_\lambda(\mathbf{x}) = \frac{np}{\lambda} - \sum_{j=1}^n x_j$, d'où $\alpha_n^* = \frac{p}{\bar{X}}$, avec $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$.

ii) On suppose que $X_1, \ldots, X_n$ sont indépendantes suivent une loi $\mathcal{N}(m, \sigma^2)$, de densité f_{m,σ^2}. On a, avec $\theta = (m, \sigma^2)$,

$$f_\theta(\mathbf{x}) = \sigma^{-n} (2\pi)^{-n/2} \exp -\frac{1}{2\sigma^2} \sum_{j=1}^{n} (x_j - m)^2,$$

$$\ln f_\theta(\mathbf{x}) = -\frac{n}{2} \ln 2\pi - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{j=1}^{n} (x_j - m)^2,$$

$$\frac{d}{dm} \ln f_\theta(\mathbf{x}) = -\frac{1}{\sigma^2} \sum_{j=1}^{n} (x_j - m), \ \frac{d}{d\sigma} \ln f_\theta(\mathbf{x}) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{j=1}^{n} (x_j - m)^2,$$

d'où

$$m^* = \frac{1}{n} \sum_{j=1}^{n} X_j = \bar{X}, \ (\sigma^2)^* = \frac{1}{n} \sum_{j=1}^{n} (X_j - \bar{X})^2 = S^2.$$

iii) Lorsque $X_1 \sim \mathcal{B}(1,p)$ sa densité est

$$f_p(x) = p^x(1-p)^{1-x}.$$

On a

$$\ln f_p(\mathbf{x}) = n[\overline{x}\ln p + (1-\overline{x})\ln(1-p)],$$

d'où $\frac{d}{dp}\ln f_p(\mathbf{x}) = \frac{\overline{x}}{p} - \frac{1-\overline{x}}{1-p}$ et donc $p^* = \overline{X}$.

iv) Supposons que $X_1 \sim \mathcal{U}([\theta, 1+\theta])$. Si $X_1, \ldots, X_n$ sont indépendantes de même loi. Alors

$$f_\theta(\mathbf{X}) = \begin{cases} 1 & \text{si } \theta \leqslant X_{(1)} \leqslant \ldots \leqslant X_{(n)} \leqslant 1+\theta \\ 0 & \text{sinon }, \end{cases}$$

où $X_{(1)} \leqslant \ldots \leqslant X_{(n)}$ est la suite $X_1, \ldots, X_n$ réordonnée par ordre croissant (statistiques d'ordre). On peut prendre

$$\hat{\theta}_n^* = X_{(1)} \text{ ou } \hat{\theta}_n^* = X_{(n)} - 1.$$

v) Supposons que $X_1 \sim \mathcal{U}([0,\theta])$, $\theta > 0$, de densité

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} & \text{si } x \in [0,\theta] \\ 0 & \text{sinon }. \end{cases}$$

Alors

$$\hat{\theta}_n^*(\mathbf{X}) = \max_{j=1,\ldots,n} X_j = X_{(n)}.$$

Proposition 1.1 *Soit $q : \Theta \to \Lambda$ un changement de paramètre bijectif. Si $\hat{\theta}^*$ est un estimateur de maximum de vraisemblance du paramètre θ, alors $q(\hat{\theta}^*)$ sera un estimateur de maximum de vraisemblance du paramètre $\lambda = q(\theta)$ pour la famille paramétrique $\mathcal{Q} = \{Q_\lambda = P_{q^{-1}(\lambda)} : \lambda \in \Lambda\}$ dominée par μ.*

Preuve. On note g_λ la densité de Q_λ par rapport à μ. Pour tout borélien $B \in \mathcal{B}(E)$ on peut écrire :

$$Q_\lambda(B) = P_{q^{-1}(\lambda)}(B) = \int_B f_{qi^{-1}(\lambda)}(x)\mu(dx)$$

et

$$\mathrm{P}_\theta(B) = Q_{q(\theta)}(B) = \int_B g_{q(\theta)}(x)\mu(dx).$$

On en déduit que μ-p.p. $g_\lambda = f_{q^{-1}(\lambda)}$ et $f_\theta = g_{q(\theta)}$. Ainsi μ-p.p.

$$g_{\hat{\lambda}^*}(x) = \sup_{\lambda\in\Lambda} g_\lambda(x) = \sup_{\lambda\in\Lambda} f_{q^{-1}(\lambda)}(x) \leqslant f_{\hat{\theta}^*}(x)$$

et

$$f_{\hat{\theta}^*}(x) = \sup_{\theta\in\Theta} f_\theta(x) = \sup_{\theta\in\Theta} g_{q(\theta)}(x) \leqslant g_{\hat{\lambda}^*}(x).$$

$\square$

REMARQUE. On peut écrire l'estimateur de maximum de vraisemblance comme une fonctionnelle de la distribution empirique. En effet, soient $X_1, \ldots, X_n$ variables aléatoires indépendantes de même loi P_θ. On note $\hat{\theta}_n^*$ l'estimateur de maximum de vraisemblance. On a :

$$\frac{1}{n}\sum_{i=1}^{n} \ln f_{\hat{\theta}^*}(X_i) = \sup_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} \ln f_\theta(X_i) = \sup_{\theta\in\Theta} \int \ln f_\theta(x)\hat{Q}_n(dx),$$

avec $\hat{Q}_n$ la distribution empirique. Ainsi

$$\hat{\theta}^* = \text{argsup} \int \ln f_\theta(x)\hat{Q}_n(dx),$$

donc une fonctionnelle de la distribution empirique.

On va donner plus bas les résultats généraux de convergence de la suite des estimateurs de maximum de vraisemblance ainsi que de normalité asymptotique. Donnons deux idées de preuve simples sous des hypothèses plus fortes que nécessaire.

Définition 1.6 *Soient deux lois de probabilité* $\mathrm{P}_1, \mathrm{P}_2$ *sur* $\mathcal{B}(E)$ *de densités* f_1, f_2 *par rapport à la mesure* μ. **L'information de Kullback-Leibler** *de* P_2 *par rapport à* P_1 *est donnée par*

$$H(f_2 \mid f_1) := \int_E f_2(x) \ln \frac{f_2(x)}{f_1(x)} \mu(dx),$$

lorsque f_2 *n'est pas strictement positive au points* x *où* $f_1(x) = 0$ *; lorsque* $f_2(x) > 0$ *si* $f_1(x) = 0$, *on posera* $H(\cdot \mid \cdot) = +\infty$.

REMARQUE : On peut voir que

$$H(f_2 \mid f_1) = \int_E \frac{f_2(x)}{f_1(x)} \ln \frac{f_2(x)}{f_1(x)} f_1(x)\mu(dx)$$

$$= \int_E \left[\frac{f_2(x)}{f_1(x)} \ln \frac{f_2(x)}{f_1(x)} - \frac{f_2(x)}{f_1(x)} + 1\right] f_1(x)\mu(dx) = \int_E \chi(\frac{f_2(x)}{f_1(x)}) f_1(x)\mu(dx),$$

où $\chi(u) := u \ln u - u + 1 \geqslant 0$, $\forall u > 0$, et $\chi(u) > 0$ si $u \neq 1$. Ainsi $H(f_2 \mid f_1) \geqslant 0$ avec égalité si et seulement $f_2 = f_1$, μ-p.p.

REMARQUE : On notera $H(\theta_2 \mid \theta_1) = H(f_{\theta_2} \mid f_{\theta_1})$ et on supposera que H est bien définie pour tous $\theta_1, \theta_2 \in \Theta$, que $H \geqslant 0$ (finie), que $H = 0$ seulement sur la diagonale $\theta_2 = \theta_1$ (c'est l'hypothèse (A_0)).

Théorème 1.7 *Supposons que les hypothèses* (A_0) *et* (A_μ) *sont satisfaites et que :*

a) $\theta \mapsto \ln f_\theta(x)$ *est continue en* θ *pour tout* $x \in E$ *;*

b) Θ est un compact de $\mathbb{R}^k$.

Alors

$$\hat{\theta}_n^* \overset{\mathrm{P}_\theta - p.s.}{\longrightarrow} \theta, \forall \theta \in \Theta.$$

Idée de preuve. Fixons $\theta_0 \in \Theta$. On va montrer

$$\hat{\theta}_n^* \overset{P_{\theta_0} - \text{p.s.}}{\longrightarrow} \theta_0.$$

Notons

$$\mathsf{e}(\theta) := \mathrm{E}_{\theta_0}\left[\ln f_\theta(X_1)\right].$$

Alors, $\mathsf{e}(\theta_0) - \mathsf{e}(\theta) = H(\theta_0 \mid \theta) \geqslant 0$, par la remarque précédente, donc

$$\mathsf{e}(\theta) \leqslant \mathsf{e}(\theta_0), \forall \theta \in \Theta, \text{ avec égalité si et seulement si } \theta = \theta_0.$$

Par la loi des grands nombres on sait que, pour tout $\theta \in \Theta$,

$$\frac{1}{n}\sum_{j=1}^{n} \ln f_\theta(X_j) \overset{P_{\theta_0} - \text{p.s.}}{\longrightarrow} \mathrm{E}_{\theta_0}\left[\ln f_\theta(X_1)\right] = \mathsf{e}(\theta).$$

Notons Ω' l'ensemble de probabilité un sur lequel cette convergence a lieu. L'idée est que $\mathsf{e}(\theta)$ a un unique maximum en $\theta = \theta_0$, les fonctions approchantes ont leur maxima globaux proches de θ_0 avec probabilité proche de 1, d'où la conclusion. En effet, on sait que

$$\frac{1}{n}\sum_{j=1}^{n} \sup_{\theta \in \Theta} \ln f_\theta(X_j) = \frac{1}{n}\sum_{j=1}^{n} \ln f_{\hat{\theta}_n^*}(X_j),$$

donc, en particulier, pour tout $\omega \in \Omega'$

$$(*) \qquad \frac{1}{n}\sum_{j=1}^{n} \ln f_{\theta_0}(X_j(\omega)) \leqslant \frac{1}{n}\sum_{j=1}^{n} \ln f_{\hat{\theta}_n^*(\omega)}(X_j(\omega)).$$

et le membre de gauche tend vers $\mathsf{e}(\theta_0) = \max_{\theta \in \theta} \mathsf{e}(\theta)$. Supposons que $\hat{\theta}_n^*(\omega)$ ne converge pas vers θ_0. Comme Θ est un compact, quitte à renommer une sous-suite, on peut dire que, pour tout $\omega \in \Omega'$

$$\hat{\theta}_n^*(\omega) \to \theta_1 \in \Theta, \, \theta_1 \neq \theta_0.$$

De plus, on peut choisir $\eta > 0$ tel que

$$\mathsf{e}(\theta_1) \leqslant \mathsf{e}(\theta_0) - \eta.$$

Par la continuité, pour tout $\omega \in \Omega'$,

$$\ln f_{\hat{\theta}_n^*(\omega)}(X_j(\omega)) \to \ln f_{\theta_1}(X_j(\omega)),$$

donc, pour tout $\varepsilon > 0$, il existe $n_0(\varepsilon, \omega)$ tel que si $n \geqslant n_0$

$$|\ln f_{\hat{\theta}_n^*(\omega)}(X_j(\omega)) - \ln f_{\theta_1}(X_j(\omega))| < \varepsilon.$$

On écrit, pour $n \geqslant n_0$,

$$\left| \frac{1}{n} \sum_{j=1}^{n} \ln f_{\hat{\theta}_n^*(\omega)}(X_j(\omega) - \frac{1}{n} \sum_{j=1}^{n} \ln f_{\theta_1}(X_j(\omega)) \right|$$

$$\leqslant \frac{1}{n} \sum_{j=1}^{n_0-1} |\ln f_{\hat{\theta}_n^*(\omega)}(X_j(\omega)) - \ln f_{\theta_1}(X_j(\omega))| + \frac{1}{n} \sum_{j=n_0}^{n} |\ln f_{\hat{\theta}_n^*(\omega)}(X_j(\omega)) - \ln f_{\theta_1}(X_j(\omega))|$$

$$\leqslant \frac{1}{n} \sum_{j=1}^{n_0-1} \text{cst.} + \frac{1}{n} n\varepsilon \leqslant 2\varepsilon$$

encore par la continuité de $\ln f_\theta(x)$ sur le compact Θ. On en déduit que, pour tout $\omega \in \Omega'$,

$$\frac{1}{n} \sum_{j=1}^{n} \ln f_{\hat{\theta}_n^*(\omega)}(X_j(\omega) - \frac{1}{n} \sum_{j=1}^{n} \ln f_{\theta_1}(X_j(\omega)) \to 0$$

tandis que

$$\frac{1}{n} \sum_{j=1}^{n} \ln f_{\theta_1}(X_j(\omega)) \to \mathsf{e}(\theta_1).$$

En revenant à la relation (*) on obtient, pour tout $\omega \in \Omega'$,

$$\mathsf{e}(\theta_0) \leftarrow \frac{1}{n} \sum_{j=1}^{n} \ln f_{\theta_0}(X_j(\omega)) \leqslant \frac{1}{n} \sum_{j=1}^{n} \ln f_{\hat{\theta}_n^*(\omega)}(X_j(\omega) \to \mathsf{e}(\theta_1) \leqslant \mathsf{e}(\theta_0) - \eta$$

avec $\eta > 0$; absurde. □

Remarque : En pratique, Θ est tout $\mathbb{R}^k$ ou une partie non-compacte. On peut supposer

a') $\theta \mapsto \ln f_\theta(x)$ est continue pour tout $x \in E$;

b') il existe une famille croissante $\{\Theta_t : t > 0\}$ de compacts de $\mathbb{R}^k$ telle que $\Theta_t \uparrow \Theta$ et pour tout $\theta_0 \in \Theta$,

$$\lim_{t\to\infty} \mathrm{E}_{\theta_0} \left[\sup_{\theta \in \Theta_t^c} \ln f_\theta(X_1) \right] = -\infty \text{ et } \mathrm{E}_{\theta_0} \left[\sup_{\theta \in \Theta_t} |\ln f_\theta(X_1)| \right] < \infty, \forall t > 0.$$

En effet, soit la suite de fonctions aléatoires sur Θ,

$$\theta \mapsto \frac{1}{n} \sum_{j=1}^{n} \ln f_\theta(X_j) =: \Xi_n(\theta)$$

qui converge sur chaque Θ_t vers $\mathsf{e}(\theta)$ qui a un unique maximum en θ_0. Il faut éviter que le maximum soit dans Θ_t^c pour t grand, alors e.m.v. est proche de θ_0. Cela veut dire qu'avec probabilité qui tend vers 1,

$$\sup_{\theta\in\Theta_t^c} \Xi_n(\theta) < \mathsf{e}(\theta_0) - 1,$$

pour t fixé, mais grand. Par ailleurs, par la loi des grands nombres,

$$\sup_{\theta\in\Theta_t^c} \Xi_n(\theta) \leqslant \frac{1}{n}\sum_{j=1}^n \sup_{\theta\in\Theta_t^c} \ln f_\theta(X_j) \to \mathrm{E}_{\theta_0}\left[\sup_{\theta\in\Theta_t^c} \ln f_\theta(X_1)\right]$$

qui est proche de $-\infty$ si t est grand.
On peut tester (exercice) que la famille $\{\gamma(p,1) : p > 0\}$ satisfait les hypothèses a'), b').□

Passons à l'étude de la normalité asymptotique. On a besoin d'une hypothèse en plus de (A_0) et (A_μ) :

(A_1) les distributions de la famille $\mathcal{P}$ ont toutes le même support,
c'est-à-dire $\{x : f_\theta(x) > 0\}$ ne dépend pas de θ.

Définition 1.7 *Sous* (A_0), (A_1) *et* (A_μ), **l'information de Fisher** *est définie par*

$$I(\theta) := \mathrm{E}_\theta\left[\left(\frac{d}{d\theta}\ln f_\theta(X_1)\right)^2\right],\ \theta\in\Theta.$$

REMARQUE : On peut voir que :

$$\mathrm{E}_\theta\left[\frac{d}{d\theta}\ln f_\theta(X_1)\right] = \int_E \frac{d}{d\theta}\ln f_\theta(x)\cdot f_\theta(x)\mu(dx) = \int_E \frac{d}{d\theta} f_\theta(x)\mu(dx) = \frac{d}{d\theta}\int_E f_\theta(x)\mu(dx) = 0.$$

Aussi,

$$\frac{d^2}{d\theta^2}\ln f_\theta(x) = -\frac{1}{f_\theta(x)^2}\left(\frac{d}{d\theta}f_\theta(x)\right)^2 + \frac{1}{f_\theta(x)}\frac{d^2}{d\theta^2}f_\theta(x).$$

Alors

$$\mathrm{E}_\theta\left[\frac{d^2}{d\theta^2}\ln f_\theta(X_1)\right] = -\mathrm{E}_\theta\left[\frac{1}{f_\theta(X_1)^2}\left(\frac{d}{d\theta}f_\theta(X_1)\right)^2\right] + \mathrm{E}_\theta\left[\frac{1}{f_\theta(X_1)}\frac{d^2}{d\theta^2}f_\theta(X_1)\right]$$

$$= -\int_E \frac{1}{f_\theta(x)}\left(\frac{d}{d\theta}f_\theta(x)\right)^2\mu(dx) + \frac{d^2}{d\theta^2}\int_E f_\theta(x)\mu(dx) = -\int_E\left(\frac{d}{d\theta}\ln f_\theta(x)\right)^2\cdot f_\theta(x)\mu(dx)$$

$$= -\mathrm{E}_\theta\left[\left(\frac{d}{d\theta}\ln f_\theta(X_1)\right)^2\right] = -\mathrm{Var}_\theta\left[\frac{d}{d\theta}\ln f_\theta(X_1)\right] = -I(\theta).$$

On peut voir qu'il y a une relation entre l'information de Kullback-Leibler et l'information de Fisher :

$$\frac{d^2}{d\vartheta^2} H(\theta \mid \vartheta)_{|\vartheta=\theta} = I(\theta).$$

□

Théorème 1.8 *Supposons que les hypothèses* (A_0), (A_μ) *et* (A_1) *sont satisfaites et que :*

c) $\theta \mapsto \ln f_\theta(x)$ *est de classe* C^2 *et on a* $I(\theta) > 0$, *pour tout* $\theta \in \Theta \subset \mathbb{R}$ *;*

d) le maximum de la vraisemblance est atteint dans un point intérieur de Θ *et tel que* $\hat{\theta}_n^*$ *est solution de l'équation de vraisemblance :*

$$\sum_{j=1}^{n} \frac{d}{d\theta} \ln f_\theta(x_j) = 0.$$

Alors

$$\sqrt{n}(\hat{\theta}_n^* - \theta) \overset{\mathrm{P}_\theta - loi}{\longrightarrow} \mathcal{N}(0, \frac{1}{I(\theta)}).$$

Idée de preuve. Fixons à nouveau $\theta_0 \in \Theta$. On montre que

$$\sqrt{n}(\hat{\theta}_n^* - \theta_0) \overset{P_{\theta_0}-\text{loi}}{\longrightarrow} \mathcal{N}(0, \frac{1}{I(\theta_0)})$$

Notons

$$\Xi_n(\theta, x_1, \ldots, x_n) := \sum_{j=1}^{n} \frac{d}{d\theta} \ln f_\theta(x_j).$$

Alors on peut écrire

$$\Xi_n(\hat{\theta}_n^*, X_1, \ldots, X_n) - \Xi_n(\theta_0, X_1, \ldots, X_n) = (\hat{\theta}_n^* - \theta_0) \sum_{j=1}^{n} \frac{d^2}{d\theta^2} \ln f_{\breve{\theta}_n}(X_j),$$

où $\breve{\theta}_n$ est situé entre $\hat{\theta}_n^*$ et θ_0. Par la convergence de $\hat{\theta}_n^*$ vers θ_0 on déduit

$$\breve{\theta}_n \overset{P_{\theta_0}-\text{p.s.}}{\longrightarrow} \theta_0.$$

On écrit

$$(**) \qquad -\frac{1}{\sqrt{n}} \Xi_n(\theta_0, X_1, \ldots, X_n) = \left[\sqrt{n}(\hat{\theta}_n^* - \theta_0)\right] \cdot \left[\frac{1}{n} \sum_{j=1}^{n} \frac{d^2}{d\theta^2} \ln f_{\breve{\theta}_n}(X_j)\right]$$

Par le théorème central limite et par la remarque précédente

$$\frac{1}{\sqrt{n}} \Xi_n(\theta_0, X_1, \ldots, X_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{d}{d\theta} \ln f_{\theta_0}(X_j) \overset{P_{\theta_0}-\text{loi}}{\longrightarrow} \mathcal{N}(0, I(\theta_0))$$

Par la continuité, pour tout $\omega \in \Omega'$,

$$\frac{d^2}{d\theta^2} \ln f_{\breve{\theta}_n(\omega)}(X_j(\omega)) - \frac{d^2}{d\theta^2} \ln f_{\theta_0}(X_j(\omega)) \to 0,$$

d'où, comme dans la preuve du théorème précédent, pour tout $\omega \in \Omega'$,

$$\frac{1}{n}\sum_{j=1}^{n} \frac{d^2}{d\theta^2} \ln f_{\breve{\theta}_n(\omega)}(X_j(\omega)) - \frac{1}{n}\sum_{j=1}^{n} \frac{d^2}{d\theta^2} \ln f_{\theta_0}(X_j(\omega)) \to 0.$$

Enfin, par la loi des grands nombres, pour tout $\omega \in \Omega'$,

$$\frac{1}{n}\sum_{j=1}^{n} \frac{d^2}{d\theta^2} \ln f_{\theta_0}(X_j(\omega)) \to I(\theta_0)$$

encore par la remarque précédente, d'où,

$$\frac{1}{n}\sum_{j=1}^{n} \frac{d^2}{d\theta^2} \ln f_{\breve{\theta}_n}(X_j) \overset{P_{\theta_0}-\text{p.s.}}{\longrightarrow} I(\theta_0).$$

En remplaçant en (**), on trouve

$$\sqrt{n}(\hat{\theta}_n^* - \theta_0) \overset{P_{\theta_0}-\text{loi}}{\longrightarrow} \mathcal{N}(0, \frac{1}{I(\theta_0)}).$$

$\square$

Pour énoncer les résultats dans leur plus grande généralité, on va faire encore une hypothèse :

($\mathrm{A_O}$) $\quad \Theta$ est un intervalle ouvert (pas nécessairement fini) $]\underline{\theta}, \bar{\theta}[$

On admettra le résultat suivant qui donne les propriétés asymptotiques de l'estimateur de maximum de vraisemblance. θ_0 est un point quelconque de Θ.

Théorème 1.9 *Supposons satisfaites les conditions (A_0), (A_μ), (A_1) et (A_O).*

1) (**convergence**) *Supposons que pour tout x la densité $f_\theta(x)$ est dérivable par rapport à θ sur Θ, avec la dérivée $f'_\theta(x)$. Soit l'équation de vraisemblance*

$$f'_\theta(\mathbf{x}) = \ln f_\theta(\mathbf{x}) = 0.$$

Alors avec une probabilité qui tend vers 1, l'équation de vraisemblance admet une racine $\theta_n^* = \theta_n^*(\mathbf{x})$ telle que, lorsque $n \to \infty$,

$$\theta_n^*(\mathbf{X}) \overset{prob}{\longrightarrow} \theta_0.$$

De plus si l'équation de vraisemblance admet une racine unique θ_n^*, pour chaque n et tous $\mathbf{x}$, alors avec probabilité qui tend vers 1, $\hat{\theta}_n^* = \theta_n^*(\mathbf{X})$ est l'estimateur de maximum de vraisemblance.

*2) (***normalité asymptotique***) Supposons que pour tout x la densité $f_\theta(x)$ est trois fois dérivable par rapport à θ sur Θ, avec dérivée troisième continue et bornée par une fonction intégrable : pour tout x et tout θ*

$$|\frac{d^3}{d\theta^3}\ln f_\theta(\mathbf{x})| \leqslant M(\mathbf{x}), \quad \textit{avec } \mathrm{E}_{\theta_0}[M(\mathbf{X})] < \infty.$$

Alors toute suite convergente $\theta_n^* = \theta_n^*(\mathbf{x})$ racines de l'équation de vraisemblance satisfait

$$\sqrt{n}(\theta_n^*(\mathbf{X}) - \theta_0) \xrightarrow{loi} \mathcal{N}(0, \frac{1}{I(\theta_0)}),$$

où $I(\theta)$ est l'information de Fisher.

1.5 Comparaison des estimateurs. Efficacité. Inégalité de Cramer-Rao

Soit $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\}$ une famille paramétrique de probabilités. Soient $X_1, \ldots, X_n$ variables aléatoires indépendantes de même loi $\mathrm{P}_\theta \in \mathcal{P}$.

1.5.1 Paramètre scalaire

On supposera dans ce paragraphe qu'il s'agit du cas scalaire : $\theta \in \Theta \subset \mathbb{R}$.

Définition 1.8 *Un estimateur θ^* est dit* **sans biais** *si, pour tout $\theta \in \Theta$, $\mathrm{E}_\theta(\theta^*) = \theta$. On appelle* **biais** *la fonction $b(\theta) = \mathrm{E}_\theta(\theta^*) - \theta$. Un estimateur avec un biais non nul est dit* **biaisé***.*

REMARQUE : On pourra examiner (exercice) quels estimateurs trouvés précédemment sont ou non sans biais.

Exemple. Supposons que le n-échantillon est de loi $\mathcal{N}(m, \sigma^2)$. Alors la moyenne empirique $\bar{X}$ est un estimateur sans biais de m. La variance empirique S^2 est un estimateur biaisé de σ^2, tandis que

$$S_0^2 := \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

est un estimateur sans biais de σ^2.

Définition 1.9 *Pour un estimateur θ^* on définit* **la qualité "quadratique"** *de θ^*,*

$$d^2(\theta) = \mathrm{E}_\theta[(\theta^* - \theta)^2].$$

Un estimateur θ_1^* *est* **meilleur** *que* θ_2^* *si*

$$\mathrm{E}_\theta[(\theta_1^* - \theta)^2] =: d_1^2(\theta) \leqslant d_2^2(\theta) := \mathrm{E}_\theta[(\theta_2^* - \theta)^2]$$

pour tout $\theta \in \Theta$ *et s'il existe un* $\theta' \in \Theta$ *tel que* $d_1^2(\theta') < d_2^2(\theta')$. *Un estimateur* θ^* *est dit* **admissible** *s'il n'existe pas d'estimateur meilleur que* θ^*.

Remarque : En général, il n'existe pas d'estimateur meilleur que tous les autres. Soit $\theta_1 \in \Theta$ et $\theta_1^* = \theta_1 = \text{cst.}$; alors $\mathrm{E}_\theta[(\theta_1^* - \theta)^2] = 0$ pour $\theta = \theta_1$. Ainsi, si θ^* est un estimateur meilleur que tous les autres, alors $d^2(\theta_1) = \mathrm{E}_{\theta_1}[(\theta^* - \theta_1)^2] = 0$. Comme θ_1 est arbitraire, il vient $d^2(\theta) = 0$ pour tout θ, et donc $\theta^* = \theta$ p.s., pour tout θ, ce qui est absurde.

Définition 1.10 *Un estimateur* $\theta_0^* \in K$ *est* **efficace dans la classe** K *(qui est une partie stricte de l'ensemble des estimateurs) si*

$$d_0^2(\theta) = \mathrm{E}_\theta[(\theta_0^* - \theta)^2] \leqslant \mathrm{E}_\theta[(\theta^* - \theta)^2] = d^2(\theta),\ \forall \theta^* \in K,\ \forall \theta \in \Theta.$$

On désignera par K_0 **l'ensemble des estimateurs sans biais**. *Un estimateur efficace dans* K_0 *sera dit simplement* **efficace**. *Il est donc* **sans biais, de variance minimale**.

Remarque : Un estimateur sans biais n'existe pas nécessairement. Si θ^* est un tel estimateur, avec $\theta^* = g(X)$, on doit avoir $\int g(x) \mathrm{P}_\theta(dx) = \theta$, soit $\int g(x) f_\theta(x) \mu(dx) = \theta$. Prenons $X \sim \mathcal{B}(n, p)$ et posons $\theta = \varphi(p)$, où φ est une fonction donnée. On a donc

$$\sum_{k=1}^{n} C_n^k g(k) p^k (1-p)^{n-k} = \varphi(p).$$

Le membre de gauche de cette équation est un polynôme de degré plus petit ou égal à n en p, ce qui n'est pas nécessairement φ.

Théorème 1.10 *Soit* $b : \Theta \to \mathbb{R}$ *et on note* $K_b = \{\theta^* : \mathrm{E}_\theta[\theta^*] - \theta = b(\theta)\}$ *la classe des estimateurs de biais fixé* b. *Il existe au plus un estimateur efficace dans* K_b. *Ceci est en particulier vrai avec* $b(\theta) \equiv 0$, *c'est-à-dire dans la classe* K_0 *des estimateurs sans biais.*

Preuve. Soient θ_0^* et θ_1^* deux estimateurs efficaces dans K_b. Pour $i = 0,\ 1$, posons

$$\Delta_i = \theta_i^* - \theta,\ D = \mathrm{Var}_\theta(\theta_i^*) = \mathrm{E}_\theta\left[(\theta_i^* - \theta - b(\theta))^2\right],\ \theta^* = \frac{\theta_0^* + \theta_1^*}{2}.$$

Comme θ_0^* et θ_1^* sont efficaces dans K_b, on peut écrire

$$d_0^2(\theta) \leqslant d_1^2(\theta) \ \text{ et } \ d_1^2(\theta) \leqslant d_0^2(\theta)\, \forall \theta \in \Theta,$$

donc, $d_0^2 = d_1^2$. Il est clair que $D_i = \mathrm{E}_\theta(\Delta_i^2) - b(\theta)^2 = d_i^2(\theta) - b(\theta)^2$ donc $D_i := D$ ne dépend pas de i. Puisque :

$$\left(\frac{\Delta_0 + \Delta_1}{2}\right)^2 + \left(\frac{\Delta_0 - \Delta_1}{2}\right)^2 = \frac{\Delta_0^2 + \Delta_1^2}{2},$$

et comme $\frac{\Delta_0+\Delta_1}{2} = \theta^* - \theta$, $\Delta_0 - \Delta_1 = \theta_0^* - \theta_1^*$, on a

$$\mathrm{E}_\theta(\theta^* - \theta)^2 + \frac{1}{4}\mathrm{E}_\theta(\theta_0^* - \theta_1^*)^2 = D + b^2(\theta)$$

Mais $\theta^* \in K_b$, puisque θ_0^* et $\theta_1^* \in K_b$. D'où :

$$\mathrm{E}_\theta(\theta^* - \theta)^2 = d^2(\theta) \geqslant d_i^2(\theta) = D + b^2(\theta).$$

Donc

$$\mathrm{E}_\theta(\theta_0^* - \theta_1^*)^2 \leqslant 0$$

c'est-à-dire $\theta_0^* = \theta_1^*$ p.s. □

Le théorème suivant répond à la question suivante : jusqu'où peut-on descendre avec $d^2(\theta)$?

Théorème 1.11 *(inégalité de Cramer-Rao) Supposons que les hypothèses* (A_0)*,* (A_1) *et* (A_μ) *sont satisfaites et que :*

a) $\theta \mapsto \ln f_\theta(x)$ *est de classe* C^1 *pour tout* x *et telle que l'information de Fisher* $I(\theta) > 0$, $\forall \theta \in \Theta \subset \mathbb{R}$.

b) $\theta \mapsto b(\theta)$ *est dérivable et* $\theta^* \in K_b$ *est une estimateur tel que* $\mathrm{E}_\theta[(\theta^*)^2] < \infty$

Alors

$$\mathrm{Var}_\theta(\theta^*) \geqslant \frac{[1 + b'(\theta)]^2}{nI(\theta)}, \ \text{ donc } \ d^2(\theta) \geqslant \frac{[1 + b'(\theta)]^2}{nI(\theta)} + b(\theta)^2, \ \forall \theta \in \Theta.$$

En particulier, si $\theta^* \in K_0$ *est un estimateur sans biais,*

$$\mathrm{Var}_\theta(\theta^*) = d^2(\theta) \geqslant \frac{1}{nI(\theta)}, \ \forall \theta \in \Theta.$$

Preuve. On a $\mathrm{E}_\theta[\theta^*] = \theta + b(\theta)$, donc $\frac{d}{d\theta}\mathrm{E}_\theta[\theta^*] = 1 + b'(\theta)$. Mais

$$\frac{d}{d\theta}\mathrm{E}_\theta[\theta^*] = \int_{E^n} \theta^*(x_1, \ldots, x_n)\frac{d}{d\theta} f_\theta(x_1, \ldots, x_n)\mu(dx_1)\ldots\mu(dx_n)$$

$$= \mathrm{E}_\theta\left[\theta^* \frac{d}{d\theta}\ln f_\theta(X_1, \ldots, X_n)\right] = \mathrm{E}_\theta\left[(\theta^* - \theta - b(\theta))\frac{d}{d\theta}\ln f_\theta(X_1, \ldots, X_n)\right],$$

car $\mathrm{E}_\theta\left[\frac{d}{d\theta}\ln f_\theta(\mathbf{X})\right] = 0$. Ainsi, par l'inégalité de Cauchy-Schwarz,

$$(1 + b'(\theta))^2 = \left\{\frac{d}{d\theta}\mathrm{E}_\theta[\theta^*]\right\}^2 = \left\{\mathrm{E}_\theta\left[(\theta^* - \theta - b(\theta))\frac{d}{d\theta}\ln f_\theta(X_1, \ldots, X_n)\right]\right\}^2$$

$$\leqslant \mathrm{E}_\theta\left[(\theta^* - \theta - b(\theta))^2\right] \cdot \mathrm{E}_\theta\left[\left(\frac{d}{d\theta}\ln f_\theta(X_1, \ldots, X_n)\right)^2\right] = \mathrm{Var}_\theta(\theta^*) \cdot nI(\theta),$$

puisque, l'information de Fisher d'un n-échantillon est

$$\mathrm{E}_\theta\left[\left(\frac{d}{d\theta}\ln f_\theta(X_1, \ldots, X_n)\right)^2\right] = \mathrm{Var}_\theta\left[\sum_{i=1}^n \frac{d}{d\theta}\ln f_\theta(X_i)\right] = \sum_{i=1}^n \mathrm{Var}_\theta\left[\frac{d}{d\theta}\ln f_\theta(X_i)\right].$$

□

Corollaire 1.2 *Sous les conditions du théorème, si pour $\theta_0^* \in K_b$ l'inégalité de Cramer-Rao est vérifiée par le signe "=", alors θ_0^* est efficace dans K_b.*

Remarque : On verra que la réciproque de ce résultat est fausse : un estimateur peut être efficace sans que la borne inférieure $\frac{(1+b'(\theta))^2}{nI(\theta)}$ soit atteinte.

Définition 1.11 *Supposons remplies les conditions du théorème précédent. Un estimateur est dit R-**efficace** dans la classe K_b si :*

$$d^2(\theta) = \mathrm{E}_\theta[(\theta^* - \theta)^2] = \mathrm{Var}_\theta(\theta^*) + b^2(\theta) = \frac{(1 + b'(\theta))^2}{nI(\theta)} + b^2(\theta),\ \forall \theta \in \Theta,$$

autrement dit, si l'égalité est réalisée dans l'inégalité de Cramer-Rao.

Exemples.
1. Soit $\mathbf{X}$ un n-échantillon de loi $\mathcal{N}(m, \sigma^2)$, avec σ^2 connu. Soit $m^* = \bar{X}$. On a

$$\ln f_m(x) = -\ln\sqrt{2\pi}\,\sigma - \frac{(x-m)^2}{2\sigma^2},\ \frac{d}{dm}\ln f_m(x) = \frac{x-m}{\sigma^2},$$

$$I(m) = \mathrm{E}_m[\frac{d}{dm}\ln f_m(X)^2] = \mathrm{E}_m\left[\frac{(X-m)^2}{\sigma^4}\right] = \frac{1}{\sigma^2}.$$

D'où

$$\mathrm{Var}_m(m^*) = \frac{1}{nI(m)} = \frac{\sigma^2}{n}.$$

Ainsi $m^* = \bar{X}$ est R-efficace.
2. Soit $\theta^* = S_1^2 = \frac{1}{n}\sum_{i=1}^n (X_i - m)^2$ un estimateur du paramètre $\theta = \sigma^2$ de la distribution gaussienne, m étant connu. On vérifie que :

$$\mathrm{Var}_\theta(\theta^*) = \mathrm{E}_\theta[(\theta^* - \sigma^2)^2] = \frac{2\sigma^4}{n},\ \frac{d}{d\theta}\ln f_\theta(x) = -\frac{1}{2\theta} + \frac{(x-m)^2}{2\theta^2},$$

$$I(\theta) = \mathrm{E}_\theta[\frac{d}{d\theta}\ln f_\theta(X)^2] = \frac{1}{4\theta^4},\ \mathrm{E}_\theta\{[(X-m)^2-\theta]^2\} = \frac{1}{2\theta^2} = \frac{1}{2\sigma^4}.$$

Ainsi $\mathrm{Var}_\theta(\theta^*) = \frac{1}{nI(\theta)}$ et l'estimateur $\theta^* = S_1^2$ est R-efficace. Notons que la variance de l'estimateur sans biais $S_0^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ est égale à $\frac{2\sigma^4}{n-1}$, si bien que cet estimateur n'est pas R-efficace, ni efficace.

Remarque : On a vu deux méthodes de construction d'estimateurs ponctuels. Les estimateurs par la méthode des moments sont souvent sans biais (ou on peut les modifier facilement pour qu'ils le soient). On a déjà discuté leur optimalité asymptotique (voir le Théorème 1.6). L'estimateur de maximum de vraisemblance, peut ne pas être sans biais de variance minimale. De même on a vu ses propriétés d'optimalité asymptotique. Qu'en est-il de l'efficacité ?

Définition 1.12 *On dit qu'un estimateur $\theta_{0,n}^* \in K$ est* **asymptotiquement efficace dans** K *si pour tout $\theta_n^* \in K$ et tout $\theta \in \Theta$:*

$$\limsup_{n\to\infty} \frac{\mathrm{E}_\theta(\theta_{0,n}^* - \theta)^2}{\mathrm{E}_\theta(\theta_n^* - \theta)^2} \leqslant 1.$$

Un estimateur θ_n^* est **asymptotiquement normal** *si, lorsque $n \to \infty$,*

$$\sqrt{n}(\theta_n^* - \theta) \xrightarrow{loi} \mathcal{N}(0, \sigma^2(\theta)).$$

On appelle $\sigma^2(\theta)$ **paramètre de dispersion** *de θ_n^*.*

Définition 1.13 *Appelons $K_\mathcal{N}$ la* **classe des estimateurs asymptotiquement normaux** *et $\mathbf{K}_{\mathcal{N},\mathbf{2}}$ la classe des estimateurs asymptotiquement normaux tels que, lorsque $n \to \infty$:*

$$\mathrm{E}_\theta\left[\sqrt{n}|\theta_n^* - \theta|\right] \to 0,\ \mathrm{E}_\theta\left[n(\theta_n^* - \theta)^2\right] \to \sigma^2(\theta) < \infty.$$

On peut voir que la restriction de $K_\mathcal{N}$ à $K_{\mathcal{N},2}$ appauvrit peu la première classe. On pourra prendre la classe $K_{\mathcal{N},2}$ sans perdre beaucoup de généralité.

Proposition 1.2 *Soit $K \subset K_{\mathcal{N},2}$. Un estimateur $\theta_{0,n}^*$ dans K est asymptotique efficace dans K si et seulement si, pour tout $\theta_n^* \in K$ et $\theta \in \Theta$,*

$$\sigma_0^2(\theta) \leqslant \sigma^2(\theta),$$

où $\sigma^2(\theta)$ et $\sigma_0^2(\theta)$ sont les paramètres de dispersion de θ_n^* et $\theta_{0,n}^*$ respectivement.

Preuve. En effet, pour tout $\theta_{0,n}^* \in K$

$$\mathrm{E}_\theta(\theta_{0,n}^* - \theta)^2 = \frac{\sigma_0^2(\theta)}{n}(1 + o(1)).$$

Alors, la définition exprime que

$$\mathrm{E}_\theta(\theta^*_{0,n}-\theta)^2 \leqslant \mathrm{E}_\theta(\theta^*_n-\theta)^2(1+o(1)),$$

pour tout $\theta^*_n \in K$. Cette inégalité est équivalente à l'inégalité de l'énoncé. □

Théorème 1.12 *Soit $K \subset K_{\mathcal{N},2}$. Si $\theta^*_{0,n}$ et $\theta^*_{1,n}$ sont asymptotiques efficaces dans K et si $\frac{1}{2}(\theta^*_{0,n}+\theta^*_{1,n}) \in K$, alors $\theta^*_{0,n}$ et $\theta^*_{1,n}$ sont asymptotiquement confondus, c'est-à-dire, lorsque $n \to \infty$,*

$$\sqrt{n}(\theta^*_{0,n}-\theta^*_{1,n}) \xrightarrow{prob} 0.$$

Remarque : On voit que dans cet énoncé le biais a disparu.

Preuve. L'assertion est une conséquence de

$$\mathrm{E}_\theta\left[(\sqrt{n}(\theta^*_0-\theta^*_1))^2\right] \to 0,$$

d'après l'inégalité Bienaymé-Tchebytchev. Posons, pour $i = 0,\, 1$,

$$M_{i,n} = \mathrm{E}_\theta\left[n(\theta^*_i-\theta)^2\right],\ \Delta_i = \theta^*_i - \theta.$$

D'où, comme dans la preuve du Théorème 1.10,

$$\mathrm{E}_\theta\left[n(\theta^*-\theta)^2\right] + \frac{1}{4}\mathrm{E}_\theta\left[n(\theta^*_0-\theta^*_1)^2\right] = \frac{M_{0,n}+M_{1,n}}{2}.$$

Par l'efficacité asymptotique de θ^*_i et la Proposition 1.13 on a $\sigma^2_0(\theta) = \sigma^2_1(\theta)$. Puisque $\theta^* \in K$, on a, encore par l'efficacité asymptotique de θ^*_i, en passant à la limite,

$$\lim_{n\to\infty} \mathrm{E}_\theta\left[n(\theta^*_0-\theta^*_1)^2\right] \leqslant 0.$$

□

Définition 1.14 *Un estimateur $\theta^*_{0,n}$ est dit* **asymptotiquement R-efficace** *si*

$$\mathrm{E}_\theta[(\theta^*_{0,n}-\theta)^2] = \frac{1+o(1)}{nI(\theta)},\ \ n \to \infty.$$

On introduit la classe K des estimateurs θ^*_n satisfaisant pour tout $\theta \in \Theta$:

$$\mathrm{E}_\theta[(\theta^*_n)^2] < \text{cst.},\ |b(\theta)| \leqslant \frac{\varepsilon(\theta,n)}{\sqrt{n}},\ |b'(\theta)| \leqslant \varepsilon(\theta,n),$$

où $\varepsilon(\theta,n) = o(1)$, lorsque $n \to \infty$.

Proposition 1.3 *Supposons remplies les conditions du théorème inégalité de Cramer-Rao. Soient $\theta^*_{0,n}$ et $\theta^*_{1,n}$ deux estimateurs asymptotiquement R-efficaces dans la classe K. Alors ils sont asymptotiquement confondus : lorsque $n \to \infty$*

$$\sqrt{n}(\theta^*_{0,n} - \theta^*_{1,n}) \xrightarrow{prob} 0.$$

Preuve. On reprend la preuve du Théorème 1.12. Puisque $\theta^*_n = (\theta^*_{0,n} + \theta^*_{1,n})/2 \in K$, en vertu de l'égalité

$$\mathrm{E}_\theta\left[n(\theta^*_n - \theta)^2\right] + \frac{1}{4}\mathrm{E}_\theta\left[n(\theta^*_0 - \theta^*_1)^2\right] = \frac{M_{0,n} + M_{1,n}}{2},\ M_{i,n} = \mathrm{E}_\theta\left[n(\theta^*_{i,n} - \theta)^2\right]$$

et de l'inégalité de Cramer-Rao, il vient

$$\limsup_{n\to\infty} \mathrm{E}_\theta\left[n(\theta^*_0 - \theta^*_1)^2\right] \leqslant 0.$$

□

1.5.2 Paramètre vectoriel

Considérons un instant le cas où $\theta = (\theta_1, \dots, \theta_k)$ et θ^* sont des vecteurs de $\mathbb{R}^k$. La notion de biais ne change pas, mais la comparaison des estimateurs est plus délicate ici car il y a seulement un ordre partiel sur $\mathbb{R}^k$.

Définition 1.15 *On dira qu'un estimateur θ^*_0 est* **efficace dans la classe** *K si pour tout vecteur $u \in \mathbb{R}^k$, l'estimateur $\alpha^*_0 =< \theta^*_0, u >$ est un estimateur efficace du paramètre $\alpha =< \theta, u >$ dans la classe des estimateurs $\alpha^* =< \theta^*, u >$, $\theta^* \in K$, c'est-à-dire que pour tous les $\theta \in \Theta$, $u \in \mathbb{R}^k$, $\theta^* \in K$,*

$$\mathrm{E}_\theta[< \theta^*_0 - \theta, u >^2] \leqslant \mathrm{E}_\theta[< \theta^* - \theta, u >^2],$$

ou, équivalent,

$$\sum_{i,j=1}^{k} \mathrm{Cov}_\theta(\theta^*_{0,i} - \theta_i, \theta^*_{0,j} - \theta_j)u_i u_j \leqslant \sum_{i,j=1}^{k} \mathrm{Cov}_\theta(\theta^*_i - \theta_i, \theta^*_j - \theta_j)u_i u_j.$$

Un estimateur efficace dans la classe des estimateurs sans biais est dit simplement **efficace***.*

Remarque : Comme on a défini l'efficacité à l'aide du cas scalaire on déduit du Théorème 1.10 qu'il existe au plus un estimateur efficace dans la classe des estimateurs efficaces $K_{\mathbf{b}}$ de biais fixé $\{\theta^* : \mathrm{E}_\theta[\theta^*] - \theta = \mathbf{b}(\theta) = (b_1(\theta), \dots, b_k(\theta))\}$.

Définition 1.16 *Sous* (A_0)*,* (A_1) *et* (A_μ)*,* **l'information de Fisher** *est définie par*

$$I(\theta) = \left(I_{j\ell}(\theta)\right)_{1\leqslant j,\ell\leqslant k}, \; I_{j\ell}(\theta) = \mathrm{E}_\theta\left[\frac{\partial}{\partial\theta_j}\ln f_\theta(X_1)\frac{\partial}{\partial\theta_\ell}\ln f_\theta(X_1)\right].$$

Remarque : On peut vérifier que :

$$\mathrm{E}_\theta\left[\frac{\partial}{\partial\theta_j}\ln f_\theta(X_1)\right] = 0,$$

$$I_{j\ell}(\theta) = \mathrm{Cov}_\theta\left(\frac{\partial}{\partial\theta_j}\ln f_\theta(X_1), \frac{\partial}{\partial\theta_\ell}\ln f_\theta(X_1)\right) = -\mathrm{E}_\theta\left[\frac{\partial^2}{\partial\theta_j\partial\theta_\ell}\ln f_\theta(X_1)\right].$$

Théorème 1.13 *(inégalité de Cramer-Rao) Supposons que* (A_0)*,* (A_1) *et* (A_μ) *sont satisfaites et que :*

a) $\theta \mapsto \ln f_\theta(x)$ *est de classe* C^1 *pour tout* x *et telle que* $I(\theta)$ *est inversible pour tout* $\theta \in \Theta \subset \mathbb{R}^k$ *;*

b) $\theta \mapsto \mathbf{b}(\theta)$ *est différentiable et* $\theta^* \in K_{\mathbf{b}}$ *tel que sa matrice de covariance existe pour tout* θ *:*

$$\mathrm{Cov}_\theta(\theta^*) = \mathrm{E}_\theta\left[(\theta^* - \theta - \mathbf{b}(\theta))^\star(\theta^* - \theta - \mathbf{b}(\theta))\right].$$

Alors

$$\mathrm{Cov}_\theta(\theta^*) \geqslant \frac{1}{n}(Id_k + B(\theta))I(\theta)^{-1}(Id_k + B(\theta))^\star,$$

où Id_k *est la matrice unité d'ordre* k *et où*

$$B(\theta) = \left(\frac{\partial b_j(\theta)}{\partial\theta_\ell}\right)_{1\leqslant j,\ell\leqslant k}.$$

Preuve. On note

$$L'_j = \sum_{i=1}^n \frac{\partial}{\partial\theta_j}\ln f_\theta(X_i), \; j = 1,\ldots,k, \; \mathbf{L}' = (L'_1,\ldots,L'_k).$$

On a $\mathrm{E}_\theta(\mathbf{L}') = 0$ et $\mathrm{E}_\theta(\theta^*)^\star\mathbf{L}' = Id_k + B(\theta)$, d'où

$$\mathrm{E}_\theta(\theta^* - \theta - \mathbf{b}(\theta))^\star\mathbf{L}' = Id_k + B(\theta).$$

On pose $\xi = (\theta^* - \theta - \mathbf{b}(\theta))^\star$ et $\eta = (\mathbf{L}')^\star$. D'après l'inégalité Cauchy-Schwarz pour les matrices :

$$\mathrm{E}_\theta(\xi\xi^\star) \geqslant \mathrm{E}_\theta(\xi\eta^\star)\left(\mathrm{E}_\theta(\eta\eta^\star)^{-1}\right)\mathrm{E}_\theta(\eta\xi^\star),$$

avec $\mathrm{E}_\theta(\xi\xi^\star) = \mathrm{Cov}_\theta(\theta^*)$, $\mathrm{E}_\theta(\eta\eta^\star) = nI(\theta)$, par indépendance des X_i et $\mathrm{E}_\theta(\xi\eta^\star) = Id_k + B(\theta)$. On trouve ainsi l'inégalité de Cramer-Rao. □

Définition 1.17 *On dit qu'un estimateur vectoriel* $\theta_0^* \in K_{\mathbf{b}}$ *est* R-**efficace** *si l'inégalité de Cramer-Rao est satisfaite par le signe "=", pour tout* θ.

Définition 1.18 *On dit qu'un estimateur vectoriel* $\theta_{1,n}^* \in K$ *est* **asymptotiquement efficace dans** K *si pour tout vecteur* $u \in \mathbb{R}^k$*, l'estimateur* $< \theta_{1,n}^*, u >$ *est un estimateur asymptotiquement efficace de* $< \theta, u >$*. Cela est équivalent à dire que pour tous* $\theta_n^* \in K$*,* $u \in \mathbb{R}^k$*,* $\theta \in \Theta$

$$\sum_{i,j=1}^{k} \sigma_{ij}^{(1)}(\theta) u_i u_j \leqslant \sum_{i,j=1}^{k} \sigma_{ij}(\theta) u_i u_j,$$

où $(\sigma_{ij}^{(1)}(\theta))_{i,j}$ *et* $(\sigma_{ij}(\theta))_{i,j}$ *sont respectivement les* **paramètres de dispersion** *de* $\theta_{1,n}^*$ *et* θ_n^**, c'est-à-dire les matrices de covariances des distributions limites de* $\sqrt{n}(\theta_{1,n}^* - \theta)$ *et* $\sqrt{n}(\theta_n^* - \theta)$*, quand* $n \to \infty$.

1.6 Statistiques exhaustives

1.6.1 Rappels sur les lois conditionnelles

Soit (X, Y) un couple de variables aléatoires. Par définition, pour toute fonction (borélienne) h, **l'espérance conditionnelle** $\mathrm{E}(h(X) \mid Y)$ est une variable aléatoire $\sigma(Y)$-mesurable. Elle est donc de la forme :

$$\mathbb{E}(h(X) \mid Y) = \mathbb{E}(h(X) \mid \sigma(Y)) = \varphi(Y) \text{ avec } \varphi \text{ borélienne.}$$

Elle est caractérisée par :

$$\forall g \text{ borélienne , } \mathbb{E}(h(X)g(Y)) = \mathbb{E}(\varphi(Y)g(Y))$$

(on sous-entend que les quantités écrites existent). Elle est donnée par un **noyau** $N(y, dx)$:

$$y \mapsto N(y, dx) \text{ est mesurable, } A \mapsto N(y, A) \text{ est une probabilité}$$

et on a

$$\mathbb{E}(h(X) \mid Y) = \int N(Y, dx) h(x) := Nh(Y).$$

Le noyau $N(Y, \cdot)$ décrit **la loi conditionnelle de** $h(X)$ **sachant** Y. Dans le cas de deux variables discrètes :

$$N(\ell \mid k) = \mathrm{P}(X = \ell \mid Y = k) \text{ et } Nh(Y) = \sum_{\ell} N(\ell \mid Y) h(\ell).$$

Dans le cas où le couple (X, Y) a une densité $p(x, y)$ par rapport à la mesure de Lebesgue :

$$n(x \mid y) = \frac{p(x, y)}{\int p(x, y) dx} \text{ et } Nh(Y) = \int n(x \mid Y) h(x) dx.$$

1.6.2 Statistiques exhaustives

Une des premières choses à faire quand on étudie les données est de réduire le nombre de données sans pour autant perdre de l'information utile. Ainsi on veut savoir s'il est possible de remplacer les données $(\mathbf{x})$ ou encore l'échantillon $\mathbf{X} = (X_1, \ldots, X_n)$ par une fonction $\mathbf{s}(\mathbf{x})$ ou, respectivement par une statistique $\mathbf{S} = \mathbf{S}(\mathbf{X})$, plus simple, mais qui contient suffisamment d'information ? L'idée est basée sur la remarque suivante : si la loi conditionnelle de $\mathbf{X}$ sachant S ne dépend pas de la loi P_θ de X, alors $\mathbf{S}$ est suffisamment informative pour $\mathcal{P}$. En effet, dans cette situation la loi conditionnelle de $\mathbf{X}$ sachant S peut être spécifiée indépendamment de $\mathcal{P}$, lorsqu'on donne $S = s$, on peut générer une variable aléatoire $\mathbf{X}'$ de même loi que $\mathbf{X}$ avec des nombres aléatoires. Donc les informations données par $\mathbf{X} = (X_1, \ldots, X_n)$ ne donnent pas plus sur $\mathcal{P}$ que $\mathbf{S}$ le fait. $\mathbf{S}$ est dite alors statistique suffisante.

Exemple : On a n pièces identiques chacune ayant la probabilité p de montrer pile. On lance successivement les n pièces et on obtient un mot de n lettres "pffp...pf" ; ensuite on lance les n pièces en une seule fois et on obtient un certain nombre k de pile. A priori la première expérience offre plus d'informations : si on a seulement le nombre de pile on ne peut pas donner la suite. Toutefois il y a C_n^k mots ayant le même nombre de pile et ils ont tous la même probabilité $p^k(1-p)^{n-k}$ de sortir. Sachant k ils ont la même probabilité $\frac{1}{C_n^k}$. Ainsi, si on sait k, l'ordre peut être choisie au hasard sans besoin de connaître p ; on ne peut pas distinguer un mot dont l'ordre à été faite au hasard du mot obtenu en lançant les pièces. Ainsi l'ordre n'est pas importante pour l'étude de p, il suffit de connaître le nombre de pile et oublier le reste.

Définition 1.19 *Soit $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\}$ un modèle paramétrique, $\Theta \subset \mathbb{R}^k$. Soient $\mathbf{X} = (X_1, \ldots, X_n)$ un n-échantillon de loi dans $\mathcal{P}$ et $\mathbf{S} = \mathbf{S}(\mathbf{X}) = (S_1(\mathbf{X}), \ldots, S_m(\mathbf{X}))$ une statistique, c'est-à-dire une variable aléatoire à valeurs dans $\mathbb{R}^m$ de l'observation. S est dite* **exhaustive** *ou* **suffisante pour** *θ si :*

$$\mathrm{E}_\theta(h(\mathbf{X}) \mid \mathbf{S}) \text{ ne dépend pas de } \theta \text{ pour toute } h.$$

En d'autres termes, la loi conditionnelle de $\mathbf{X}$ sachant $\mathbf{S}$ ne dépend pas de θ ou encore :

$$\mathrm{E}_\theta(h(X) \mid \mathbf{S}) = Nh(\mathbf{S}), \text{ où le noyau } N \text{ ne dépend pas de } \theta$$

Remarques : 1) Intuitivement, cela signifie que, conditionnellement à $\mathbf{S}$, la loi de $\mathbf{X}$ ne dépend plus de θ : toute l'information relative à θ est dans $\mathbf{S}$.
2) Dans la suite on supposera pour simplifier, que $m = k$.

Exemples. 1) Soient $\mathbf{X} = (X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{P}(\lambda)$. Soit $n\bar{X} = \sum_{i=1}^n X_i = S$. Montrons que $S = n\bar{X}$ est une statistique exhaustive. On a ici

$$N(s, x) = \mathrm{P}_\lambda(X_1 = x_1, \ldots, X_n = x_n \mid S = s) = \begin{cases} \frac{\mathrm{P}_\lambda(X_1 = x_1, \ldots, X_n = x_n)}{\mathrm{P}_\lambda(S = s)}, & \text{si } s = x_1 + \ldots + x_n \\ 0, & \text{sinon.} \end{cases}$$

Puisque $S \sim \mathcal{P}(n\lambda)$, on a, si $s = x_1 + \ldots + x_n$,

$$N(s,x) = \left(e^{-n\lambda}\frac{(n\lambda)^s}{s!}\right)^{-1}\prod_{i=1}^{n}\left(e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}\right) = \frac{s!}{n^s\prod_{i=1}^{n} x_i!}.$$

Ainsi la loi conditionnelle est la loi multinomiale $\mathcal{M}(s, \frac{1}{n}, \ldots, \frac{1}{n})$ indépendante de λ.
2) Soit un n-échantillon de loi $\mathcal{B}(1,p)$. $S = n\bar{X}$ est une statistique exhaustive car

$$\mathrm{P}_p(X_1 = x_1, \ldots, X_n = x_n \mid S = s) = \begin{cases} \frac{\mathrm{P}_p(X_1=x_1,\ldots,X_n=x_n)}{\mathrm{P}_p(S=s)}, & \text{si } s = x_1 + \ldots + x_n \\ 0, & \text{sinon.} \end{cases}$$

Mais si $s = x_1 + \ldots + x_n$,

$$\mathrm{P}_p(X_1 = x_1, \ldots, X_n = x_n \mid S = s) = \frac{\prod_{j=1}^{n} p^{x_j}(1-p)^{n-x_j}}{C_n^s p^s(1-p)^{n-s}} = \frac{1}{C_n^s},$$

indépendante de p.

Exemple. Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un n-échantillon de loi dans $\{\mathcal{N}(m,1) : m \in \mathbb{R}\}$. Donc

$$\mathbf{X} \sim \mathcal{N}_n(m\mathbf{1}, Id_n),$$

où $\mathbf{1}$ est le vecteur ayant toutes ses coordonnées égales à 1 et où Id_n est la matrice identité. Montrons que la moyenne empirique $\bar{X} = \frac{1}{n}\mathbf{1}^\star\mathbf{X}$ est une statistique suffisante. Soit H une matrice orthogonale dont on sait que la dernière ligne est $\frac{1}{\sqrt{n}}\mathbf{1}^\star$. Le vecteur aléatoire

$$\mathbf{Y} = H\mathbf{X} \sim \mathcal{N}_n(mH\mathbf{1}, Id_n), \text{ où } H\mathbf{1} = (0, \ldots, 0, \sqrt{n})^\star.$$

Ainsi $Y_1, \ldots, Y_{n-1}$ sont indépendantes de même loi $\mathcal{N}(0,1)$ et indépendantes de $Y_n \sim \mathcal{N}(\sqrt{n}m, 1)$. D'où la loi conditionnelle de $(Y_1, \ldots, Y_{n-1})$ sachant $Y_n = \sqrt{n}\bar{X}$ est $\mathcal{N}_{n-1}(\mathbf{0}, Id_{n-1})$ et donc la loi conditionnelle de toute fonction borélienne de $\mathbf{X} = (X_1, \ldots, X_n)$ sachant $\bar{X}$ est indépendante de m.

Théorème 1.14 *(de factorisation de Neyman-Fisher)*
Supposons satisfaite la condition (A_μ). Pour que $\mathbf{S}$ *soit exhaustive, il faut et il suffit que $f_\theta(\mathbf{x}) = \prod_{i=1}^{n} f_\theta(x_i)$ soit de la forme*

$$(*) \quad f_\theta(\mathbf{x}) = g(\mathbf{S}(\mathbf{x}), \theta)h(\mathbf{x}) \ \ \mu^n - p.p.$$

où g et h sont deux fonctions boréliennes positives, g dépend de $\mathbf{x} = (x_1, \ldots, x_n)$ par l'intermédiaire de $\mathbf{S}$ *et h ne dépend pas de θ.*

Remarque : Cette représentation n'est pas unique.

Exemple. Revenons à l'exemple de la loi de Poisson :

$$f_\lambda(\mathbf{x}) = e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}.$$

Soit $s = \sum_{i=1}^n x_i$; f_λ est de la forme précédente avec

$$g(s,\lambda) = e^{-n\lambda}\lambda^s,\ h(\mathbf{x}) = \prod_{i=1}^n \frac{1}{x_i!}.$$

Exemple. Soit $X \sim \mathcal{N}(m,\sigma^2)$, $\theta = (m,\sigma^2)$.

$$\begin{aligned} f_\theta(\mathbf{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-m)^2}{2\sigma^2}} &= \sigma^{-n}(2\pi)^{-\frac{n}{2}} \exp - \left(\sum_{i=1}^n \frac{(x_i-m)^2}{2\sigma^2}\right) \\ &= \sigma^{-n} \exp - \left(\frac{\sum_{i=1}^n x_i^2 - 2mn\bar{x} + nm^2}{2\sigma^2}\right)(2\pi)^{-\frac{n}{2}}. \end{aligned}$$

Posons $S_1 = n\bar{x}$, $S_2 = \sum_{i=1}^n x_i^2$. On a, avec $\mathbf{S} = (S_1,S_2)$, $f_\theta(\mathbf{x}) = g(S(\mathbf{x}),\theta)h(x)$ avec

$$g(\mathbf{S},\theta) = \sigma^{-n} \exp - \frac{S_2 - 2mS_1 + m^2}{2\sigma^2},\ h(\mathbf{x}) = (2\pi)^{-\frac{n}{2}}.$$

Le couple (S_1,S_2) est donc exhaustif.

Exemple. Soit $X \sim \mathcal{U}_{[0,\theta]}$; alors

$$f_\theta(\mathbf{x}) = \begin{cases} \frac{1}{\theta^n}, & \text{si } x_i \in [0,\theta],\ \forall i = 1,\ldots,n \\ 0, & \text{sinon .} \end{cases}$$

Définissons $x_{(1)} = \min_{i=1,\ldots,n} x_i$, $x_{(n)} = \min_{i=1,\ldots,n} x_i$. On a donc $f_\theta(\mathbf{x}) = g(x_{(n)},\theta)h(\mathbf{x})$, avec

$$h(\mathbf{x}) = \begin{cases} 1, & \text{si } x_{(1)} \geqslant 0 \\ 0, & \text{sinon ,} \end{cases} \quad g(s,\theta) = \begin{cases} \theta^{-n}, & \text{si } s \leqslant \theta \\ 0, & \text{sinon ,} \end{cases}$$

En d'autres termes $S = X_{(n)}$ est une statistique exhaustive, où $X_{(n)}$ est la n-ième statistique d'ordre.

Preuve du Théorème 1.14.
i) (cas discret) Les valeurs possibles de $\mathbf{X}$ sont alors $\{\mathbf{x}^1, \mathbf{x}^2, \ldots\}$. Nous allons noter (pour simplifier l'écriture) $\mathbf{s}^j = \mathbf{S}(\mathbf{x}^j)$ et on va montrer que $\mathrm{P}_\theta(\mathbf{X} = \mathbf{x}^k \mid \mathbf{S} = \mathbf{s}^j)$, $\forall j,k$, ne dépend pas de θ. On va vérifier cette propriété sur $\Theta_j := \{\theta : \mathrm{P}_\theta(\mathbf{S} = \mathbf{s}^j) > 0\}$. Supposons (*) réalisée. On a

$$\mathrm{P}_\theta(\mathbf{S} = \mathbf{s}^j) = \sum_{\mathbf{x}:\mathbf{S}(\mathbf{x})=\mathbf{s}^j} \mathrm{P}_\theta(\mathbf{X} = \mathbf{x}).$$

$$= \sum_{\mathbf{x}:\mathbf{S}(\mathbf{x})=\mathbf{s}^j} f_\theta(\mathbf{x}) = \sum_{\mathbf{x}:\mathbf{S}(\mathbf{x})=\mathbf{s}^j} g(\mathbf{s}^j,\theta)h(\mathbf{x}) = g(\mathbf{s}^j,\theta) \sum_{\mathbf{x}:\mathbf{S}(\mathbf{x})=\mathbf{s}^j} h(\mathbf{x}),$$

d'où

$$\mathrm{P}_\theta(\mathbf{X} = \mathbf{x}^k \mid \mathbf{S} = \mathbf{s}^j) = \frac{\mathrm{P}_\theta(\mathbf{X} = \mathbf{x}^k, \mathbf{S} = \mathbf{s}^j)}{\mathrm{P}_\theta(\mathbf{S} = \mathbf{s}^j)} = \begin{cases} \frac{\mathrm{P}_\theta(\mathbf{X}=\mathbf{x}^k)}{\mathrm{P}_\theta(\mathbf{S}=\mathbf{s}^j)} & \text{si } \mathbf{S}(\mathbf{x}^k) = \mathbf{s}^j \\ 0 & \text{sinon.} \end{cases}$$

$$= \begin{cases} \frac{f_\theta(\mathbf{x})}{\mathrm{P}_\theta(\mathbf{S}=\mathbf{s}^j)} & \text{si } \mathbf{S}(\mathbf{x}^k) = \mathbf{s}^j \\ 0 & \text{sinon.} \end{cases} = \begin{cases} \frac{g(\mathbf{s}^j,\theta)h(\mathbf{x}^k)}{g(\mathbf{s}^j,\theta)\sum_{\mathbf{x}:\mathbf{S}(\mathbf{x})=\mathbf{s}^j} h(\mathbf{x})} & \text{si } \mathbf{S}(\mathbf{x}^k) = \mathbf{s}^j \\ 0 & \text{sinon.} \end{cases}$$

$$= \begin{cases} \frac{h(\mathbf{x}^k)}{\sum_{\mathbf{x}:\mathbf{S}(\mathbf{x})=\mathbf{s}^j} h(\mathbf{x})} & \text{si } \mathbf{S}(\mathbf{x}^k) = \mathbf{s}^j \\ 0 & \text{sinon.} \end{cases}$$

qui ne dépend pas de θ.

Soit maintenant $\mathbf{S}$ une statistique suffisante. On pose

$$g(\mathbf{s}^j, \theta) = \mathrm{P}_\theta(\mathbf{S} = \mathbf{s}^j) \text{ et } h(\mathbf{x}^j) = \mathrm{P}_\theta(\mathbf{X} = \mathbf{x}^j \mid \mathbf{S} = \mathbf{S}(\mathbf{x}^j)),\ j \geqslant 1.$$

Alors, il est clair que

$$f_\theta(\mathbf{x}^j) = \mathrm{P}_\theta(\mathbf{X} = \mathbf{x}^j) = \mathrm{P}_\theta(\mathbf{X} = \mathbf{x}^j \mid \mathbf{S} = \mathbf{S}(\mathbf{x}^j))\mathrm{P}_\theta(\mathbf{S} = \mathbf{S}(\mathbf{x}^j)) = h(\mathbf{x}^j)g(\mathbf{S}(\mathbf{x}^j), \theta)\ j \geqslant 1.$$

ii) (cas à densité) Faisons la preuve dans le cas où μ est la mesure de Lebesgue. Le cas où μ est plus générale est similaire. De plus, on va faire quelques hypothèses simplificatrices, mais le résultat est vrai dans toute sa généralité. On suppose $k \leqslant n$ et qu'il existe $S_j = S_j(\mathbf{X})$, $j = k+1, \ldots, n$, statistiques, donc $\mathbf{S}' = \mathbf{S}'(\mathbf{X}) = (S_{k+1}(\mathbf{X}), \ldots, S_n(\mathbf{X}))$, telle que l'application

$$s_j = S_j(\mathbf{x}),\ j = 1, \ldots, n$$

est un difféomorphisme sur $\mathbb{R}^n$, d'inverse

$$x_j = X_j(\mathbf{s}, \mathbf{s}'),\ \mathbf{s}' = (s_{k+1}, \ldots, s_n),\ j = 1, \ldots, n$$

de jacobien $J(\mathbf{s}, \mathbf{s}') \neq 0$ (indépendante de θ). On note ce difféomorphisme F, donc $\mathbf{s}, \mathbf{s}') = F(\mathbf{x})$. Supposons (*) réalisée. Alors

$$f_{\mathbf{S},\mathbf{S}'}(\mathbf{s}, \mathbf{s}'; \theta) = f_\theta(F^{-1}(\mathbf{s}, \mathbf{s}'))|J(\mathbf{s}, \mathbf{s}')| = g(\mathbf{s}, \theta)h(\mathbf{x}(\mathbf{s}, \mathbf{s}'))|J(\mathbf{s}, \mathbf{s}')|$$

$$= g(\mathbf{s}, \theta)\bar{h}(\mathbf{s}, \mathbf{s}'),$$

avec $\bar{h}(\mathbf{s}, \mathbf{s}') = h(\mathbf{x}(\mathbf{s}, \mathbf{s}'))|J(\mathbf{s}, \mathbf{s}')|$, d'où

$$f_{\mathbf{S}}(\mathbf{s}; \theta) = \int_{\mathbb{R}^{n-k}} f_{\mathbf{S},\mathbf{S}'}(\mathbf{s}, \mathbf{s}'; \theta) d\mathbf{s}' = \int_{\mathbb{R}^{n-k}} g(\mathbf{s}, \theta)\bar{h}(\mathbf{s}, \mathbf{s}') d\mathbf{s}' = g(s, \theta)\bar{\bar{h}}(\mathbf{s}).$$

On déduit

$$f_{\mathbf{S}'|\mathbf{S}}(\mathbf{s} \mid \mathbf{s}') = \frac{f_{\mathbf{S},\mathbf{S}'}(\mathbf{s}, \mathbf{s}'; \theta)}{f_{\mathbf{S}}(\mathbf{s}; \theta)} = \frac{g(\mathbf{s}, \theta)\bar{h}(\mathbf{s}, \mathbf{s}')}{g(s, \theta)\bar{\bar{h}}(\mathbf{s})} = \frac{\bar{h}(\mathbf{s}, \mathbf{s}')}{\bar{\bar{h}}(\mathbf{s})}$$

qui ne dépend pas de θ. Ainsi la loi conditionnelle de $\mathbf{S}'$ sachant $\mathbf{S}$ ne dépend pas de θ, donc la loi conditionnelle de $(\mathbf{S}, \mathbf{S}')$ sachant $\mathbf{S}$ ne dépend pas de θ. Comme il y a une bijection (qui ne dépend pas de θ) entre $(\mathbf{S}, \mathbf{S}')$ et $\mathbf{X}$, $\mathbf{X} = F^{-1}(\mathbf{S}, \mathbf{S}')$, on conclut que la loi conditionnelle de $\mathbf{X}$ sachant $\mathbf{S}$ ne dépend pas de θ.

Réciproquement, supposons que $\mathbf{S}$ est une statistique exhaustive. Il est alors clair que la loi conditionnelle de $\mathbf{S}'$ sachant $\mathbf{S}$ ne dépend pas de θ. (en utilisant le même difféomorphisme). Alors

$$f_\theta(\mathbf{x}) = f_{\mathbf{S},\mathbf{S}'}(F(\mathbf{x});\theta)|J^{-1}(\mathbf{x})| = f_{\mathbf{S},\mathbf{S}'}(\mathbf{S}(\mathbf{x}),\mathbf{S}'(\mathbf{x});\theta)|J^{-1}(\mathbf{x})|$$

$$= f_{\mathbf{S}'|\mathbf{S}}(\mathbf{S}(\mathbf{x}) \mid \mathbf{S}'(\mathbf{x}))|J^{-1}(\mathbf{x})| f_{\mathbf{S}}(\mathbf{S}(\mathbf{x});\theta) = h(\mathbf{x})g(\mathbf{S}(\mathbf{x}),\theta),$$

c'est-à-dire (*). □

Corollaire 1.3 *Soient $\varphi : \mathbb{R}^k \to \mathbb{R}^k$ et $q : \mathbb{R}^k \to \mathbb{R}^k$ deux applications bijectives boréliennes bimesurables. Si $\mathbf{S} = \mathbf{S}(\mathbf{X}) \in \mathbb{R}^k$ est une statistique exhaustive pour $\theta \in \Theta \subset \mathbb{R}^k$, alors $\mathbf{S}' = \varphi(\mathbf{S})$ est aussi exhaustive pour θ et $\mathbf{S}$ est exhaustive pour $\lambda = q(\theta)$.*

Preuve. On a $\mathbf{S} = \varphi^{-1}(\mathbf{S}')$, d'où

$$f_\theta(\mathbf{x}) = g(\mathbf{S}(\mathbf{x}),\theta)h(\mathbf{x}) = g(\varphi^{-1}(\mathbf{S}'(\mathbf{x})),\theta)h(\mathbf{x}),$$

ce qui prouve que $\mathbf{S}'$ est exhaustive pour θ. Ensuite $\theta = q^{-1}(\lambda)$, d'où l'égalité

$$f_\theta(\mathbf{x}) = g(\mathbf{S}(\mathbf{x}),\theta)h(\mathbf{x})$$

devient

$$f'_\lambda(\mathbf{x}) = g'(\mathbf{S}(\mathbf{x}),\lambda)h(\mathbf{x}),$$

avec les notations $f'_\lambda(\mathbf{x}) := f_{q^{-1}(\lambda)}(\mathbf{x})$ et $g'(\mathbf{S}(\mathbf{x}),\lambda) := g(\mathbf{S}(\mathbf{x}),q^{-1}(\lambda))$. Ainsi $\mathbf{S}$ est exhaustive pour θ'. □

Corollaire 1.4 *Soit $\mathbf{X}$ un n-échantillon de loi appartenant à une famille paramétrique $\mathcal{P}$. On suppose satisfaites les hypothèses (A_0) et (A_μ). Soit $\mathbf{S} = \mathbf{S}(\mathbf{X})$ une statistique exhaustive pour θ. Alors, si $\hat{\theta}_n^*$ est l'unique estimateur du maximum de vraisemblance, il ne dépend que de $\mathbf{S}$.*

Preuve. On a $f_\theta(\mathbf{x}) = g(\mathbf{S}(\mathbf{x}),\theta)h(\mathbf{x})$. Alors $\sup_{\theta\in\Theta} f_\theta(\mathbf{x}) = \sup_{\theta\in\Theta} g(\mathbf{S}(\mathbf{x}),\theta)$, d'où :

$$\hat{\theta}_n^* = \arg\sup f_\theta(\mathbf{X}) = \arg\sup g(\mathbf{S}(\mathbf{X}),\theta).$$

□

1.6.3 Statistique exhaustive minimale

Il existe beaucoup de statistiques exhaustives. Par exemple $\mathbf{S}(X) = X$ est évidemment exhaustive. Elle est dite triviale. Lorsqu'on trouve une statistique exhaustive on peut se demander si elle est la plus "économique".

Définition 1.20 *Une statistique* $\mathbf{S}_1$ *est* **subordonnée** *à* $\mathbf{S}_2$ *si* $\mathbf{S}_1$ *est une fonction borélienne de* $\mathbf{S}_2$*,* $\mathbf{S}_1 = \varphi(\mathbf{S}_2)$*. Intuitivement, cela signifie que* $\mathbf{S}_1$ *est plus "économique" que* $\mathbf{S}_2$*.* $\mathbf{S}_1$ *et* $\mathbf{S}_2$ *sont dites* **équivalents** *si* $\mathbf{S}_1$ *est subordonnée à* $\mathbf{S}_2$ *et* $\mathbf{S}_2$ *à* $\mathbf{S}_1$*).* $\mathbf{S}_1$ *et* $\mathbf{S}_2$ *sont équivalents si et seulement si* $\mathbf{S}_1 = \varphi(\mathbf{S}_2)$*, avec* φ *bijective et bimesurable.*

Définition 1.21 *Une statistique exhaustive* $\mathbf{S}_0$ *est dite* **minimale** *si elle est subordonnée à toute statistique exhaustive* $\mathbf{S}$*.*

REMARQUE. On pourrait rendre cette notion plus intuitive en utilisant les tribus. Une tribu $\mathcal{S}$ est dite **exhaustive** pour θ si l'espérance conditionnelle $\mathrm{E}_\theta(h(X) \mid \mathcal{S})$ ne dépend pas de θ (ou si la loi conditionnelle de $h(X)$ sachant $\mathcal{S}$ ne dépend pas de θ). Ainsi $\mathbf{S}$ est **exhaustive** pour θ si la tribu engendrée par $\mathbf{S}$, $\sigma(\mathbf{S})$, est exhaustive pour θ. La notion de subordination est alors simple : $\mathbf{S}_1$ est **subordonnée** à $\mathbf{S}_2$ si $\sigma(\mathbf{S}_1) \subset \sigma(\mathbf{S}_2)$. Donc $\mathbf{S}_1$ est plus "économique" que $\mathbf{S}_2$ si la tribu $\sigma(\mathbf{S}_1)$ est plus pauvre que $\sigma(\mathbf{S}_2)$. Ainsi la **tribu exhaustive minimale** est la tribu exhaustive qui s'immerge dans toute tribu exhaustive.

Théorème 1.15 *Soit* $\mathbf{X}$ *un* n*-échantillon de loi dans une famille paramétrique dont le paramètre est scalaire* ($\theta \in \Theta \subset \mathbb{R}$) *et on suppose satisfaites les hypothèses* (A_0) *et* (A_μ)*. On suppose que la fonction vraisemblance* $\theta \mapsto f_\theta(\mathbf{x})$ *est continue (à droite ou à gauche) pour tout* $\mathbf{x}$*. Si l'estimateur du maximum de vraisemblance* $\hat{\theta}_n^*$ *est unique et est une statistique exhaustive, alors c'est une statistique exhaustive minimale.*

Preuve. Soit S une statistique exhaustive quelconque. Il s'agit de voir que $\hat{\theta}_n^*$ est $\sigma(\mathbf{S})$-mesurable : on aura alors $\hat{\theta}_n^* = \varphi(S)$ (c'est à dire que $\hat{\theta}_n^*$ est subordonnée à S). Par le théorème de factorisation :

$$f_\theta(\mathbf{x}) = g(S(\mathbf{x}), \theta)h(\mathbf{x}), \ \mu^n - p.p.$$

En fait, on peut s'arranger pour que la relation ci-dessus soit vraie partout : on modifie $f_\theta(\mathbf{x})$ en $\mathbf{x}$ sur un ensemble dense dénombrable (de θ) : cela ne modifie pas P_θ, et on utilise la continuité à droite (en θ). Mais (voir aussi le Corollaire 1.3) le point de maximum absolu de $f_\theta(\mathbf{x})$ est aussi point de maximum absolu de $g(S(\mathbf{x}), \theta)$. Puisque $\hat{\theta}_n^*$ est unique, on a pour tout t :

$$\{\hat{\theta}_n^*(\mathbf{X}) < t\} = \left\{ \sup_{\theta < t} g(S(\mathbf{X}), \theta) > \sup_{\theta \geqslant t} g(S(\mathbf{X}), \theta) \right\}.$$

Mais, par continuité à droite en θ :

$$\sup_{\theta < t} g(S(\mathbf{x}), \theta) = \sup_{\theta < t, \theta \in \mathbb{Q}} g(S(\mathbf{x}), \theta)$$

et de même

$$\sup_{\theta \geqslant t} g(S(\mathbf{x}), \theta) = \sup_{\theta \geqslant t, \theta \in \mathbb{Q}} g(S(\mathbf{x}), \theta).$$

Ceci prouve que :

$$\{\hat{\theta}_n^*(\mathbf{X}) < t\} = \left\{ \sup_{\theta < t, \theta \in \mathbb{Q}} g(S(\mathbf{X}), \theta) > \sup_{\theta \geqslant t, \theta \in \mathbb{Q}} g(S(\mathbf{X}), \theta) \right\}.$$

Ainsi $\hat{\theta}_n^*$ est mesurable par rapport à $\sigma(S)$. □

Théorème 1.16 *Soit* $\mathbf{S}$ *une statistique telle que l'équivalence suivante a lieu :*

$$\mathbf{S}(\mathbf{x}) = \mathbf{S}(\mathbf{y}) \Leftrightarrow \text{ le rapport } \frac{f_\theta(\mathbf{x})}{f_\theta(\mathbf{y})} \text{ ne dépend pas de } \theta.$$

Alors $\mathbf{S}$ *est une statistique exhaustive minimale.*

Preuve. On va faire la preuve dans le cadre scalaire : $\Theta \subset \mathbb{R}$. Montrons que S est exhaustive. Soit $z \in \mathbb{R}$ et on va fixer $\mathbf{x}_z \in S^{-1}(\{z\})$. Alors, pour tout $\mathbf{x}$, $S(\mathbf{x}) = S(\mathbf{x}_{S(\mathbf{x})})$. Alors le rapport

$$\frac{f_\theta(\mathbf{x})}{f_\theta(\mathbf{y})} = \frac{f_\theta(\mathbf{x})}{f_\theta(\mathbf{x}_{S(\mathbf{x})})}$$

ne dépend pas de θ. On note ce rapport par $h(\mathbf{x})$ et alors on a le critère de factorisation avec $g(S(\mathbf{x}), \theta) = f_\theta(\mathbf{x}_{S(\mathbf{x})})$.
Montrons que S est minimale. Soit S' une autre statistique exhaustive. Alors

$$f_\theta(\mathbf{y}) = g'(S'(\mathbf{y}), \theta) h'(\mathbf{y}),\ \forall \mathbf{y}.$$

Soient $\mathbf{x}, \mathbf{y}$ tels que $S'(\mathbf{x}) = S'(\mathbf{y})$. Alors le rapport

$$\frac{f_\theta(\mathbf{x})}{f_\theta(\mathbf{y})} = \frac{g'(S'(\mathbf{x}), \theta) h'(\mathbf{x})}{g'(S'(\mathbf{y}), \theta) h'(\mathbf{y})} = \frac{h'(\mathbf{x})}{h'(\mathbf{y})}$$

ne dépend pas de θ, donc $S(\mathbf{x}) = S(\mathbf{y})$, d'où S est une fonction de S'. □

1.7 Construction d'estimateurs efficaces

1.7.1 Améliorer un estimateur

Voici une façon d'améliorer un estimateur :

Théorème 1.17 *(Rao-Blackwell)*
Soit $\mathbf{S} = \mathbf{S}(\mathbf{X})$ *une statistique exhaustive et* θ^* *un estimateur dans* $K_{\mathbf{b}}$*, c'est-à-dire de biais égal à* $\mathbf{b}$*. Soit :*

$$\theta^*_{\mathbf{S}} = \mathrm{E}_\theta(\theta^* \mid \mathbf{S}).$$

Alors :

1. *$\theta^*_{\mathbf{S}}$ ne dépend pas de θ, et dépend de $\mathbf{X}$ par l'intermédiaire de $\mathbf{S}$;*
2. *$\theta^*_{\mathbf{S}} \in K_b$, c'est-à-dire $\mathrm{E}_\theta(\theta^*_{\mathbf{S}}) = \theta + \mathbf{b}(\theta)$, $\forall \theta \in \Theta$;*
3. *dans le cas où $\Theta \subset \mathbb{R}$, $d^2_S(\theta) := \mathrm{E}_\theta[(\theta^*_S - \theta)^2] \leqslant \mathrm{E}_\theta[(\theta^* - \theta)^2] = d^2(\theta)$, $\forall \theta \in \Theta$, avec égalité pour $\theta^*_S = \theta^*$, P_θ-p.s.*

Remarque : On peut énoncer une propriété comme dans le point *3* du théorème pour le cadre vectoriel.

Preuve. *1.* On a, puisque $\mathbf{S}$ est exhaustive :

$$\theta^*_{\mathbf{S}} = \mathrm{E}_\theta(\theta^*(\mathbf{X}) \mid \mathbf{S}) = N\theta^*(\mathbf{S}),$$

ce qui prouve que $\theta^*_{\mathbf{S}}$ ne dépend pas de θ (N ne dépend pas de θ car $\mathbf{S}$ est exhaustive) et il ne dépend de $\mathbf{X}$ que par l'intermédiaire de $\mathbf{S}$.
2.

$$\mathrm{E}_\theta(\theta^*_{\mathbf{S}}) = \mathrm{E}_\theta\left[\mathrm{E}_\theta(\theta^*(\mathbf{X}) \mid \mathbf{S})\right] = \mathrm{E}_\theta(\theta^*) = \theta + \mathbf{b}(\theta),$$

car $\theta^* \in K_{\mathbf{b}}$. En particulier, $\theta^*_{\mathbf{S}}$ est sans biais si θ^* l'est.
3. Ici $\Theta \subset \mathbb{R}$. D'après l'inégalité de Schwarz (ou Jensen) pour l'espérance conditionnelle :

$$(\theta^*_S - \theta)^2 = (\mathrm{E}_\theta(\theta^* \mid S) - \theta)^2 = [\mathrm{E}_\theta(\theta^* - \theta \mid S)]^2 \leqslant \mathrm{E}_\theta((\theta^* - \theta)^2 \mid S).$$

Prenant les espérances :

$$\mathrm{E}_\theta[(\theta^*_S - \theta)^2] \leqslant \mathrm{E}_\theta[(\theta^* - \theta)^2].$$

□

Exemple. Soient $X_1, \ldots, X_n$ variables aléatoires indépendantes de même loi de Poisson $\mathcal{P}(\lambda)$. Soit $\lambda^* = X_1$. C'est un estimateur sans biais, mais de mauvaise qualité (en particulier, puisque il ne dépend pas de n, il ne converge pas). Soit $S = n\bar{X} = \sum_{i=1}^n X_i$ une statistique exhaustive. On sait que (le vérifier) :

$$\mathrm{P}_\lambda(X_1 = k \mid S = s) = \mathrm{C}_s^k \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{s-k}, \ 0 \leqslant k \leqslant s.$$

Donc :

$$\lambda^*_S = \mathrm{E}_\lambda(X_1 \mid S) = \sum_{k=0}^{S} k\mathrm{C}_S^k \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{S-k} = \frac{S}{n} = \bar{X}$$

(la moyenne d'une loi binomiale de paramètres S et $\frac{1}{n}$).

1.7.2 Statistiques complètes

Définition 1.22 *Soit* $(Q_\theta : \theta \in \Theta)$ *une famille de probabilités sur* $(\mathbb{R}^l, \mathcal{B}(\mathbb{R}^l))$ *et on suppose que* $\theta \in \Theta \subset \mathbb{R}^k$*. On dit que la famille* $(Q_\theta : \theta \in \Theta)$ *est* **complète** *si les relations*

$$\int_{\mathbb{R}^l} \phi(s) Q_\theta(ds) = 0, \ \ \textit{pour tout} \ \ \theta \in \Theta \ \ \textit{impliquent} \ \ \phi \equiv 0, \, Q_\theta - p.p.$$

$(\phi : \mathbb{R}^l \to \mathbb{R}^k)$*. Une statistique* $\mathbf{S}$ *est* **complète** *si la famille de ses lois sous* P_θ, $\theta \in \Theta$*, est complète. En d'autres termes,* $\mathbf{S}$ *est* **complète** *si*

$$\mathrm{E}_\theta(\phi(\mathbf{S})) = 0, \, \forall \theta \in \Theta \ \Rightarrow \ \phi \equiv 0, \, \mathrm{P}_\theta - p.p.$$

Exemple. Soient $X_1, \ldots, X_n$ variables aléatoires indépendantes de même loi de Poisson $\mathcal{P}(\lambda)$. Alors $S = n\bar{X} = \sum_{i=1}^n X_i$ est complète. On a en effet :

$$\sum_{k \geqslant 0} \phi(k) e^{-n\lambda} \frac{(n\lambda)^k}{k!} = 0, \, \forall \lambda > 0$$

ou

$$\sum_{k \geqslant 0} \phi(k) \frac{z^k}{k!} = 0, \ \ \text{pour} \ \ |z| < 1,$$

ce qui prouve que la fonction analytique $\sum_{k \geqslant 0} \phi(k) \frac{z^k}{k!} = 0$, donc $\phi(k) = 0$, pour tout $k \in \mathbb{N}$.

Exemple. Soient $X_1, \ldots, X_n$ variables aléatoires indépendantes de même loi $\mathcal{U}_{[0,\theta]}$. La statistique $S = X_{(n)} = \sup_{1 \leqslant i \leqslant n} X_i$ est complète. En effet :

$$\mathrm{P}(S < s) = \left(\frac{s}{\theta}\right)^n, \ \ \text{pour tout} \ \ s \in [0, \theta].$$

La deuxième équation devient

$$\int_0^\theta \phi(s) n \frac{s^{n-1}}{\theta^n} ds = 0, \ \ \text{pour tout} \ \ \theta > 0,$$

c'est-à-dire,

$$\int_0^\theta \phi(s) s^{n-1} ds = 0,$$

ce qui implique $\phi(s) = 0$ p.p.

Théorème 1.18 *Une condition nécessaire et suffisante pour qu'une statistique* $\mathbf{S}$ *soit complète est que, pour un* $\mathbf{b}_0(\theta)$*, il existe un seul élément* θ^* *dans la classe de tous les estimateurs* $\sigma(\mathbf{S})$*-mesurables de* $K_{\mathbf{b}_0}$*. S'il existe un seul estimateur* $\sigma(\mathbf{S})$*-mesurable dans* $K_{\mathbf{b}_0}$*, il en est de même dans toute autre classe* $K_{\mathbf{b}}$*.*

Preuve. Supposons qu'il existe dans $K_{\mathbf{b}_0}$ deux estimateurs $\sigma(\mathbf{S})$-mesurables, $\theta_1^* = \varphi_1(\mathbf{S})$ et $\theta_2^* = \varphi_2(\mathbf{S})$. On a :

$$\mathrm{E}_\theta(\varphi_1(\mathbf{S})) = \mathrm{E}_\theta(\theta_1^*) = \mathrm{E}_\theta(\theta_2^*) = \mathrm{E}_\theta(\varphi_2(\mathbf{S})) = \mathbf{b}_0(\theta)$$

et donc

$$\mathrm{E}_\theta\left((\varphi_1 - \varphi_2)(\mathbf{S})\right) = 0$$

d'où, comme $\mathbf{S}$ est complète, $\varphi_1(\mathbf{s}) = \varphi_2(\mathbf{s})$ P_θ-p.p. Ainsi $\theta_1^* = \theta_2^*$ p.s.

Réciproquement, soit $\mathrm{E}_\theta(\phi(\mathbf{S})) = 0$ pour tout $\theta \in \Theta$. Si $\theta^* = \varphi(\mathbf{S}) \in K_{\mathbf{b}}$, est un estimateur dans $K_{\mathbf{b}}$ $\sigma(\mathbf{S})$-mesurable, alors l'estimateur $\sigma(\mathbf{S})$-mesurable

$$\theta_1^* = \theta^* + \phi(\mathbf{S}) = \varphi(\mathbf{S}) + \phi(\mathbf{S})$$

est dans $K_{\mathbf{b}}$ et donc, par unicité $\phi \equiv 0$ p.p. □

Théorème 1.19 *(d'unicité de Lehmann-Scheffé)*
Si une statistique exhaustive $\mathbf{S}$ *est complète et si* $\theta^* \in K_{\mathbf{b}}$, *alors* $\theta_{\mathbf{S}}^* = \mathrm{E}_\theta(\theta^* \mid \mathbf{S})$ *est le seul estimateur efficace de* $K_{\mathbf{b}}$.

Preuve. Soit θ^{**} un autre estimateur de $K_{\mathbf{b}}$. On construit $\theta_{\mathbf{S}}^{**} = \mathrm{E}_\theta(\theta^{**} \mid \mathbf{S})$; il est $\sigma(\mathbf{S})$-mesurable et $\theta_{\mathbf{S}}^{**} \in K_{\mathbf{b}}$. D'après le théorème précédent, il existe un seul estimateur $\sigma(\mathbf{S})$-mesurable dans $K_{\mathbf{b}}$, donc $\theta_{\mathbf{S}}^{**} = \theta_{\mathbf{S}}^*$. En effet $\theta_{\mathbf{S}}^*$ est aussi dans $K_{\mathbf{b}}$ et $\sigma(\mathbf{S})$-mesurable. Pour simplifier supposons qu'on est dans le cas scalaire (le cas vectoriel est identique mais l'écriture est plus compliquée). On a :

$$\mathrm{E}_\theta[(\theta_{\mathbf{S}}^* - \theta)^2] = \mathrm{E}_\theta[(\theta_{\mathbf{S}}^{**} - \theta)^2] \leqslant \mathrm{E}_\theta[(\theta^{**} - \theta)^2]$$

d'après Théorème 1.16, ou encore

$$d_{\mathbf{S}}^2(\theta) \leqslant d^2(\theta) = \mathrm{E}_\theta[(\theta^{**} - \theta)^2],\ \forall \theta \in \Theta,\ \forall \theta^{**} \in K_b.$$

L'égalité n'est possible que si $\theta^{**} = \theta_{\mathbf{S}}^*$ p.s. □

Corollaire 1.5 *Si* $\mathbf{S}$ *est une statistique exhaustive complète et* θ^* *un estimateur sans biais, alors* $\theta_{\mathbf{S}}^* = \mathrm{E}_\theta(\theta^* \mid \mathbf{S})$ *est l'unique estimateur efficace.*

Exemple. Soit X_1 de loi $\mathcal{E}(\lambda)$ donc la fonction de vraisemblance est

$$f_\lambda(\mathbf{x}) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i} \mathbb{1}_{x_i > 0}.$$

Les conditions pour l'inégalité de Cramer-Rao sont remplies dans $\Theta = \{\lambda \geqslant \delta > 0\}$. On prend $S = n\bar{X}$ une statistique exhaustive complète (le vérifier). Alors l'estimateur

$$\lambda^* = \frac{1}{\bar{X}} = \mathrm{E}_\lambda\left(\frac{1}{\bar{X}} \mid S\right)$$

est un estimateur efficace dans la classe K_b de biais $b(\lambda) = \mathrm{E}_\lambda\left(\frac{1}{\bar{X}} - \lambda\right)$. Remarquons que S est de loi $\gamma(n, \lambda)$, d'où pour $n > 1$:

$$\mathrm{E}_\lambda\left(\frac{1}{\bar{X}}\right) = n\,\mathrm{E}_\lambda\left(S^{-1}\right) = \frac{n}{n-1}\lambda.$$

L'estimateur

$$\lambda^{**} = \frac{n-1}{n\bar{X}} = \lambda^*\left(1 - \frac{1}{n}\right)$$

est donc sans biais ($n > 2$) et

$$\mathrm{E}_\lambda[(\lambda^{**})^2] = (n-1)^2\mathrm{E}_\lambda\left(S^{-2}\right) = \frac{n-1}{n-2}\lambda^2,$$

d'où

$$\mathrm{Var}_\lambda(\lambda^{**}) = \lambda^2\left(\frac{n-1}{n-2} - 1\right) = \frac{1}{n-2}\lambda^2.$$

L'estimateur λ^{**} est efficace pour $n > 2$, mais la borne inférieure n'est pas atteinte dans l'inégalité de Cramer-Rao. On a ici

$$\ln f_\lambda(\mathbf{X}) = \ln\lambda - \lambda x,\ \frac{d}{d\lambda}\ln f_\lambda(\mathbf{X}) = \frac{1}{\lambda} - x,$$

$$I(\lambda) = \mathrm{E}_\lambda[\frac{d}{d\lambda}\ln f_\lambda(\mathbf{X})^2] = \mathrm{E}_\lambda\left[\left(\frac{1}{\lambda} - X\right)^2\right] = \frac{1}{\lambda^2} - \frac{2}{\lambda^2} + \frac{2}{\lambda^2} = \frac{1}{\lambda^2}.$$

Donc, pour $n > 2$,

$$\frac{1}{nI(\lambda)} = \frac{\lambda^2}{n} < \frac{\lambda^2}{n-2} = \mathrm{Var}_\lambda(\lambda^{**}).$$

Théorème 1.20 *Toute statistique exhaustive complète* $\mathbf{S}$ *est une statistique exhaustive minimale.*

Preuve. Soit $\mathbf{S}'$ une statistique exhaustive minimale et soit $\mathcal{A}_0$ la σ-algèbre engendrée par $\mathbf{S}'$. Puisque une telle statistique $\mathbf{S}'$ est une fonction de $\mathbf{S}$, $\mathcal{A}_0 = \sigma(\mathbf{S})$. Soit :

$$\mathbf{T} := \mathbf{S} - \mathrm{E}_\theta(\mathbf{S} \mid \mathcal{A}_0).$$

$\mathbf{T}$ est en fait une fonction de $\mathbf{S}$: $\mathbf{T} := \phi(\mathbf{S})$ et on a :

$$\mathrm{E}_\theta(\mathbf{T}) = \mathrm{E}_\theta(\mathbf{S}) - \mathrm{E}_\theta(\mathrm{E}_\theta(\mathbf{S} \mid \mathcal{A}_0)) = 0,$$

soit $\mathrm{E}_\theta(\phi(\mathbf{S})) = \int \phi(\mathbf{s})d\mathrm{P}_\theta(\mathbf{s}) = 0$, pour tout $\theta \in \Theta$. Donc, comme $\mathbf{S}$ est complète, $\phi(\mathbf{s}) = 0$ P_θ-p.p. c'est-à-dire $\mathbf{S} = \mathrm{E}_\theta(\mathbf{S} \mid \mathcal{A}_0) = \mathrm{E}_\theta(\mathbf{S} \mid \mathbf{S}')$. En d'autres termes, $\mathbf{S}$ est une fonction de $\mathbf{S}'$, et donc S est minimale. $\square$

Remarque : La réciproque de ce théorème n'est pas vraie.

Exemple. Soit $X_1, \dots, X_n$ variables aléatoires indépendantes de même loi uniforme sur $[\theta, 1+\theta]$. Alors

$$f_\theta(\mathbf{x}) = \mathbb{1}_{\{\theta \leqslant x_{(1)} \leqslant x_{(n)} \leqslant 1+\theta\}},$$

si bien que $\mathbf{S} = (X_{(1)}, X_{(n)})$ est une statistique exhaustive d'après le théorème de factorisation. La densité du couple $(X_{(1)}, X_{(n)})$ est égale à :

$$g_\theta(u,v) = n(n-1)(v-u)^{n-2}\mathbb{1}_{\{\theta \leqslant u \leqslant v \leqslant 1+\theta\}}.$$

En effet :

$$\mathrm{P}(u \leqslant \inf X_i \leqslant \sup X_i \leqslant v) = \mathrm{P}(u \leqslant X_{(1)} \leqslant X_{(n)} \leqslant v) = \mathrm{P}(\cap_{i=1}^n \{u < X_i < v\})$$

$$\prod_{i=1}^n (v-u)\mathbb{1}_{\{\theta \leqslant u \leqslant v \leqslant 1+\theta\}} = (v-u)^n \mathbb{1}_{\{\theta \leqslant u \leqslant v \leqslant 1+\theta\}}$$

et

$$g_\theta(u,v) = -\frac{\partial^2}{\partial u \partial v}\mathrm{P}(u \leqslant X_{(1)} \leqslant X_{(n)} \leqslant v).$$

Montrons maintenant que $\mathbf{S}$ n'est pas complète. Supposons que :

$$0 = \int_{\mathbb{R}^2} \phi(u,v) g_\theta(u,v) du\, dv = n(n-1) \iint_{\theta<u<v<1+\theta} (v-u)^{n-2}\phi(u,v) du\, dv.$$

Choisissons $\phi(u,v) = \bar{\phi}(v-u)$ et faisons les changements de variables $v - \theta = t$, $v - u = z$:

$$0 = n(n-1)\int_0^1 dt \int_0^t z^{n-2}\bar{\phi}(z)dz = n(n-1)\int_0^1 z^{n-2}\bar{\phi}(z)dz \int_z^1 dt,$$

$$0 = n(n-1)\int_0^1 z^{n-2}\bar{\phi}(z)(1-z)dz.$$

Il est alors clair que l'on peut choisir $\bar{\phi} \neq 0$ telle que :

$$\int_0^1 z^{n-2}\bar{\phi}(z)(1-z)dz = 0.$$

Ainsi, S n'est pas complète. On peut montrer que $\mathbf{S}$ est minimale ($X_{(1)}$ et $X_{(n)}$ séparément ne sont pas suffisantes pour θ).

Le résultat suivant sera énoncé en dimension 1 mais il est vrai en dimension supérieure :

Théorème 1.21 *(de Basu)*
Soit $S = S(\mathbf{X})$ une statistique suffisante pour $\theta \in \Theta \subset \mathbb{R}$, où $\mathbf{X}$ est un n-échantillon de densité $f_\theta(\mathbf{x})$. On note $f_S(\cdot;\theta)$ la densité de S. On considère aussi $T = T(\mathbf{X})$ une autre statistique.

1. *Supposons que la densité $f_S(\cdot;\theta)$ de S a le même domaine de positivité pour tout $\theta \in \Theta$. Si T est indépendante de S alors la loi de T ne dépend pas de θ.*
2. *Supposons que S est complète. Si la loi de T ne dépend pas de θ alors T est indépendante de S.*

Preuve. *1)* On note D l'ensemble des s tels que $f_S(s;\theta) > 0$, pour tous $\theta \in \Theta$. Ainsi, par exhaustivité la densité conditionnelle de T par rapport à S $n(t \mid s)$ ne dépend pas de θ. Donc le couple (S,T) a la densité

$$f_{S,T}(s,t;\theta) = n(t \mid s) f_S(s;\theta),\ \forall t \text{ et } s \in D.$$

Mais par indépendance,

$$f_{S,T}(s,t;\theta) = f_T(t;\theta) f_S(s;\theta)\ \forall t, s \in \mathbb{R}$$

Donc

$$f_T(t;\theta) f_S(s;\theta) = n(t \mid s) f_S(s;\theta),\ \forall t \text{ et } s \in D.$$

d'où $f_T(t;\theta) = n(t \mid s)$, pour tout t et $s \in D$. On déduit que $f_T(t;\theta) = f_T(t)$ est indépendante de θ.
2) Il suffit de prouver que pour tous les $s \in \mathbb{R}$ pour lesquels $n(t \mid s)$ est définie on a $f_T(t) = n(t \mid s)$, $t \in \mathbb{R}$. Soit t fixé mais arbitraire et on considère la statistique $\phi(S;t) = f_T(t) - n(t \mid S)$, définie sauf sur un borélien N tel que $\mathrm{P}_\theta(S \in N) = 0$ pour tous les $\theta \in \Theta$. On a

$$\mathrm{E}_\theta[\phi(S;t)] = \mathrm{E}_\theta[f_T(t) - n(t \mid S)] = f_T(t) - \mathrm{E}_\theta[n(t \mid S)]$$

$$= f_T(t) - \int_{\mathbb{R}} n(t \mid s) f_S(s;\theta) ds = f_T(t) - \int_{\mathbb{R}} f_{S,T}(s,t;\theta) ds = f_T(t) - f_T(t) = 0,$$

c'est-à-dire $\mathrm{E}_\theta[\phi(S;t)] = 0$, pour tous $\theta \in \Theta$. Donc, comme S est complète, $\phi(s;t) = 0$ pour tous les $s \in N^c$ (N^c ne dépend pas de t). Ainsi $f_T(t) = n(t \mid s)$ pour $s \in N^c$. □

1.8 Familles exponentielles

On s'intéresse ici à une famille paramétrique de probabilités particulièrement importante et facile à manipuler. Le paramètre $\theta = (\theta_1, \ldots, \theta_k)$ variera dans $\Theta \subset \mathbb{R}^k$ dans cette section.

Définition 1.23 *On dit qu'une famille* $(\mathrm{P}_\theta : \theta \in \Theta)$ *est une* **famille exponentielle** *si la densité f_θ (par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ ou par rapport à la mesure de comptage d'un ensemble dénombrable) est de la forme :*

$$f_\theta(x) = h(x) \cdot \exp\left\{ \sum_{j=1}^{k} a_j(\theta) U_j(x) + V(\theta) \right\},$$

où les fonctions h, a_j, $j = 1, \ldots, k$, U_j, $j = 1, \ldots, k$, V sont finies et mesurables. On supposera, afin qu'il n'y ait pas d'ambiguïté dans la définition de f_θ que les fonctions $a_0(\theta) \equiv 1$, $a_1(\theta), \ldots, a_k(\theta)$ sont linéairement indépendantes sur Θ.

Remarque : On dit qu'il s'agit de la forme canonique (en incluant $h(x)$ dans la mesure $\mu(dx)$) si

$$f_\theta(x) = v(\theta) \exp \left\{ \sum_{j=1}^{k} \theta_j U_j(x) \right\}.$$

et le paramètre $\theta = (\theta_1, \ldots, \theta_k)$ pour lequel le membre de droite est une densité est dit **naturel**. On peut voir que l'ensemble des paramètres naturels d'une famille exponentielle canonique est un convexe. En effet, par l'inégalité de Hölder, pour tout $0 < \alpha < 1$,

$$\int \exp \left\{ \sum_{j=1}^{k} (\alpha\theta'_j + (1-\alpha)\theta''_j) U_j(x) \right\} \mu(dx)$$

$$\leqslant \left[\int \exp \left\{ \sum_{j=1}^{k} \theta'_j U_j(x) \right\} \mu(dx) \right]^{\alpha} \left[\int \exp \left\{ \sum_{j=1}^{k} \theta''_j U_j(x) \right\} \mu(dx) \right]^{1-\alpha} < \infty$$

lorsque θ' et θ'' sont tels que les intégrales sont convergentes.

Exercice : Montrons que les distributions $\mathcal{N}(m, \sigma^2)$, $\gamma_{\alpha,\lambda}$, $\mathcal{B}(1,p)$, $\mathcal{P}(\lambda)$,... sont exponentielles. Par exemple :

$$\gamma_{\alpha,\lambda}(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha} \mathbb{1}_{\{x>0\}} = \mathbb{1}_{\{x>0\}} \frac{1}{x} \exp \left\{ \lambda \ln x - \alpha x + \ln \frac{\alpha^\lambda}{\Gamma(\lambda)} \right\}.$$

$h(x) = \mathbb{1}_{\{x>0\}} \frac{1}{x}$, $a_1(\lambda, \alpha) = \lambda$, $U_1(x) = \ln x$, $a_1(\lambda, \alpha) = -\alpha$, $U_1(x) = x$, $V(\lambda, \alpha) = \ln \frac{\alpha^\lambda}{\Gamma(\lambda)}$.

Remarque : Si X_1 est de loi P_θ, avec $(\mathrm{P}_\theta : \theta \in \Theta)$ famille exponentielle, il en est de même du $(X_1, \ldots, X_n)$, où les X_i sont indépendantes et de même loi que X_1.

Théorème 1.22 *Soit $\mathbf{S} = (S_1, \ldots, S_k)$, avec $S_j = S_j(\mathbf{X}) = \sum_{i=1}^n U_j(X_i)$, $j = 1, \ldots, k$. On suppose que la fonction $\theta \mapsto \mathbf{a}(\theta)$ est telle que son image contienne un parallélépipède de $\mathbb{R}^k$. Alors $\mathbf{S}$ est une statistique exhaustive complète (et donc minimale).*

Remarque : L'hypothèse sur $\mathbf{a}$ est assez faible ; elle est remplie si Θ contient un ouvert et si la différentielle de $\mathbf{a}$ en un point de cet ouvert est de rang k, d'après le théorème des fonctions implicites.

Preuve. *i)* $\mathbf{S}$ est une statistique exhaustive. En effet

$$f_\theta(\mathbf{x}) = \prod_{i=1}^n h(x_i) \exp\left\{\sum_{j=1}^k a_j(\theta) \sum_{i=1}^n U_j(x_i) + nV(\theta)\right\}.$$

Il suffit d'appliquer le théorème de factorisation avec :

$$h(\mathbf{x}) = \prod_{i=1}^n h(x_i),\ g(\mathbf{s}, \theta) = \exp\left\{\sum_{j=1}^k a_j(\theta) s_j + nV(\theta)\right\}.$$

ii) Montrons que $\mathbf{S}$ est complète. On a

$$h(\mathbf{x}) = \prod_{i=1}^n h(x_i),\ g(\mathbf{s}, \theta) = \exp\left\{< \mathbf{a}(\theta), \mathbf{s} > + nV(\theta)\right\}.$$

Considérons sur $\mathbb{R}^k$ la mesure $\nu(B) = \int_{\mathbf{S}^{-1}(B)} h(\mathbf{x})\mu^n(d\mathbf{x})$ avec $\mathbf{S}^{-1}(B) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{S}(\mathbf{x}) \in B\}$, $B \in \mathcal{B}(\mathbb{R}^k)$. On a besoin de :

Lemme 1.4 *Soit Q_θ la loi de $\mathbf{S}(\mathbf{X})$ sous P_θ, c'est-à-dire*

$$Q_\theta(B) = \mathrm{P}_\theta(\mathbf{S}(\mathbf{X}) \in B) = \mathrm{E}_\theta[\mathbb{1}_B(\mathbf{S}(\mathbf{X}))],\ B \in \mathcal{B}(\mathbb{R}^k).$$

Alors Q_θ est absolument continue par rapport à ν et admet au point $\mathbf{s}$ la densité $g(\mathbf{s}, \theta)$.

Preuve. En effet, on a

$$Q_\theta(B) = \int_{\mathbf{S}(\mathbf{x}) \in B} g(\mathbf{S}(\mathbf{x}), \theta) h(\mathbf{x}) d\mu^n(\mathbf{x}) = \int_{\mathbf{s} \in B} g(\mathbf{s}, \theta)\nu(d\mathbf{s}),$$

après changement de variables $\mathbf{S}(\mathbf{x}) = \mathbf{s}$. □

Lemme 1.5 *Soient γ_1 et γ_2 deux mesures σ-finies positives sur $\mathbb{R}^k$. Si*

$$\int e^{<\mathbf{a},\mathbf{s}>}\gamma_1(d\mathbf{s}) = \int e^{<\mathbf{a},\mathbf{s}>}\gamma_2(d\mathbf{s}) < \infty,$$

pour tous les $\mathbf{a}$ dans un parallélépipède $I \subset \mathbb{R}^k$, alors $\gamma_1 = \gamma_2$.

Preuve. On va faire la preuve pour $k = 1$ et $I = \{a \in \mathbb{R} : |a| \leqslant \alpha\}$. Alors

$$h_j(a) = \int e^{as}\gamma_j(ds)$$

sont des fonctions analytiques pour $|a| < \alpha$. De même :

$$h_j(a+ib) = \int e^{(a+ib)s}\gamma_j(ds)$$

sont holomorphes dans la bande $|a| < \alpha$, $b \in \mathbb{R}$. Elles coïncident pour $b = 0$, donc dans la bande $|a| < \alpha$, $b \in \mathbb{R}$. En particulier, pour $a = 0$:

$$\int e^{ibs}\gamma_1(ds) = \int e^{ibs}\gamma_2(ds).$$

Mais γ_1 et γ_2 sont deux mesures positives σ-finies. L'injectivité de la transformée de Fourier permet d'en conclure que $\gamma_1 = \gamma_2$. □

On peut achever la preuve de la complétude de $\mathbf{S}$. Soit ϕ borélienne telle que :

$$\mathrm{E}_\theta(\phi(\mathbf{S})) = \int \phi(\mathbf{s})Q_\theta(d\mathbf{s}) = 0,\ \forall\theta \in \Theta.$$

On écrit $\phi = \phi^+ - \phi^-$ avec $\phi^+ = \sup(\phi, 0)$ et $\phi^- = \sup(-\phi, 0)$. On a

$$\int \phi^+(\mathbf{s})Q_\theta(d\mathbf{s}) = \int \phi^-(\mathbf{s})Q_\theta(d\mathbf{s}),$$

soit

$$\int \phi^+(\mathbf{s})g(\mathbf{s},\theta)\nu(d\mathbf{s}) = \int \phi^-(\mathbf{s})g(\mathbf{s},\theta)\nu(d\mathbf{s}),$$

$$\int \phi^+(\mathbf{s})e^{<\mathbf{a}(\theta),\mathbf{s}>}\nu(d\mathbf{s}) = \int \phi^-(\mathbf{s})e^{<\mathbf{a}(\theta),\mathbf{s}>}\nu(d\mathbf{s}),\ \forall\theta \in \Theta.$$

D'après le dernier lemme et l'hypothèse on a :

$$\phi^+(\mathbf{s})\nu(d\mathbf{s}) = \phi^-(\mathbf{s})\nu(d\mathbf{s})$$

c'est-à-dire $\phi = 0$ p.p. □

Corollaire 1.6 *Soit $\mathbf{X}$ un n-échantillon de loi dans une famille exponentielle de la forme*

$$f_\theta(x_1) = h(x_1) \cdot e^{a(\theta)U(x_1)}e^{V(\theta)},\ \theta \in \Theta \subset \mathbb{R}.$$

On note $S = S(\mathbf{X}) = \sum_{i=1}^n U(X_i)$. Si $\theta^* \in K_b$, alors $\theta^*_{\mathbf{S}} = \mathrm{E}_\theta(\theta^* \mid \mathbf{S})$ est un estimateur efficace dans K_b. En particulier, si θ^* est un estimateur sans biais, alors $\theta^*_{\mathbf{S}}$ est le seul estimateur efficace sans biais.

Le résultat suivant sera énoncé en dimension 1, mais il y a une version en dimension supérieure :

Théorème 1.23 *Soit* $\mathbf{X} = (X_1, \ldots, X_n)$ *un* n*-échantillon de loi de type exponentielle. La densité de* X *est donc*

$$f_\theta(x_1) = h(x_1) \cdot e^{a(\theta)U(x_1)} e^{V(\theta)},\ \theta \in \Theta \subset \mathbb{R}.$$

On note comme avant $S = \sum_{i=1}^n U(X_i)$. *Alors :*

1) *lorsque* S *est une variable aléatoire discrète, sa loi est toujours de la forme suivante*

$$\mathrm{P}_\theta(S = s) = \bar{h}(s) e^{a(\theta)s} e^{nV(\theta)},$$

où l'ensemble de positivité de $\bar{h}$ *est indépendante de* θ *;*

2) *lorsque* S *est à densité, sa densité est de la même forme ;*

3) *si* $T = T(\mathbf{X})$ *est une quelconque autre statistique,* T *et* S *sont indépendantes si et seulement si la loi de* T *ne dépend pas de* θ.

Preuve. On sait, par le critère de factorisation que S est une statistique suffisante pour θ.
1) On a (les sommes portent sur $\mathbf{x}$ tel que $\sum_{i=1}^n U(x_i) = s$) :

$$\begin{aligned}\mathrm{P}_\theta(S = s) &= \sum f_\theta(\mathbf{x}) = \sum \prod_{i=1}^n h(x_i) \cdot \exp\left[a(\theta) \sum_{i=1}^n U(x_i)\right] \cdot e^{nV(\theta)} \\ &= \sum \prod_{i=1}^n h(x_i) \cdot e^{a(\theta)s} e^{nV(\theta)} = \bar{h}(s) e^{a(\theta)s} e^{nV(\theta)},\end{aligned}$$

où $\bar{h}(s) = \sum \prod_{i=1}^n h(x_i)$.
2) On fait la preuve sous des hypothèses simplificatrices mais le résultat est vrai en toute généralité. On pose $Y_1 = S = \sum_{i=1}^n U(X_i)$ et $Y_i = X_i$, $i = 2, \ldots, n$. On a

$$\left\{\begin{array}{l} y_1 = \sum_{i=1}^n U(x_i) \\ y_i = x_i,\ i = 1, \ldots, n \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} U(x_1) = y_1 - \sum_{i=2}^n U(y_i) \\ y_i = x_i,\ i = 1, \ldots, n \end{array}\right.$$

ou encore

$$\left\{\begin{array}{l} x_1 = U^{-1}\left(y_1 - \sum_{i=2}^n U(y_i)\right) \\ x_i = y_i,\ i = 1, \ldots, n \end{array}\right. ,$$

où *on a suppose que* U est bijective. Ensuite

$$\frac{\partial x_1}{\partial y_1} = \frac{1}{U'(U^{-1}(z))},\ z = y_1 - \sum_{i=2}^n U(y_i),$$

à condition qu'*on suppose que* U est dérivable et *que* $U'(U^{-1}(z)) \neq 0$. Pour $i = 2, \ldots, n$ on a

$$\frac{\partial x_1}{\partial y_i} = \frac{1}{U'(U^{-1}(z))} \frac{\partial z}{\partial y_i} = -\frac{U'(y_i)}{U'(U^{-1}(z))}$$

et

$$\frac{\partial x_i}{\partial y_i} = 1,\ i = 2, \ldots, n,\ \frac{\partial x_i}{\partial y_j} = 0,\ i \neq j.$$

Alors

$$J = \frac{1}{U'(U^{-1}(z))} = \frac{1}{U'(U^{-1}(y_1 - U(y_2) - \ldots - U(y_n))}.$$

La densité de $\mathbf{Y}$ est alors

$$\begin{aligned} g_\theta(\mathbf{y}) &= e^{nV(\theta)} \exp\{a(\theta)[y_1 - U(y_2) - \ldots - U(y_n) + U(y_2) + \ldots + U(y_n)]\} \\ &\quad \times \frac{\prod_{i=2}^n h(y_i)}{h(U^{-1}[y_1 - U(y_2) - \ldots - U(y_n)])}\, |J| \\ &= e^{nV(\theta)} e^{a(\theta) y_1} \frac{\prod_{i=2}^n h(y_i)}{h(U^{-1}[y_1 - U(y_2) - \ldots - U(y_n)])}\, |J|. \end{aligned}$$

Donc si on pose

$$\bar{h}(y_1) = \iint_{\mathbb{R}^{n-1}} \frac{\prod_{i=2}^n h(y_i)}{h(U^{-1}[y_1 - U(y_2) - \ldots - U(y_n)])}\, |J|\, dy_2 \ldots dy_n,$$

on trouve le résultat désiré.
3) Il est clair que S est exhaustive et que l'ensemble de positivité de sa densité ne dépend pas de θ, par les points précédents. Par le Théorème 1.20 *1)* si T est indépendante de S alors sa loi ne dépend pas de θ. Par Théorème 1.20 *2)* si la loi de T ne dépend pas de θ, comme, par Théorème 1.21 S est complète, on en déduit que T est indépendante de S. □

1.9 Inégalité de Cramer-Rao et modèle exponentiel

On commence par étudier le cas de la dimension 1 :

Théorème 1.24 *(inégalité de Cramer-Rao et modèle exponentiel) Soit $\theta^* \in K_b$ tel que $\mathrm{E}_\theta[(\theta^*)^2] < \infty$. On rappelle que sous certaines hypothèses*

$$(*) \quad \mathrm{Var}_\theta(\theta^*) \geqslant \frac{[1 + b'(\theta)]^2}{nI(\theta)}.$$

1) Si, sur un intervalle $[\theta_1, \theta_2]$ l'égalité en (*) est réalisée et $\mathrm{Var}_\theta(\theta^*) > 0$, alors f_θ se représente sur $[\theta_1, \theta_2]$ par :

$$(**) \quad f_\theta(\mathbf{x}) = \exp[\theta^*(\mathbf{x}) \cdot C(\theta) + D(\theta)]\, h(\mathbf{x}),$$

où $C(\theta)$, $D(\theta)$ ne dépendent pas de X. En d'autres termes, la famille $\{f_\theta d\mu : \theta \in [\theta_1, \theta_2]\}$ est exponentielle.

2) Réciproquement, si $\theta^* = \text{cst.}$ ou si l'on a (**), alors l'égalité est réalisée dans (*).

Preuve du théorème. 1) L'égalité dans (*) implique que

$$\int (\theta^* - \theta - b(\theta)) f'_\theta(\mathbf{x}) \mu^n(d\mathbf{x}) = \left(\int (\theta^* - \theta - b(\theta))^2 f_\theta(\mathbf{x}) \mu^n(d\mathbf{x}) \right)^{1/2} \cdot \left(\int \frac{(f'_\theta(\mathbf{x}))^2}{f_\theta(\mathbf{x})} \mu^n(d\mathbf{x}) \right)^{1/2}.$$

D'où, puisque l'égalité dans l'inégalité de Schwarz implique la colinéarité des fonctions :

$$\frac{d}{d\theta} \ln f_\theta(\mathbf{x}) = c(\theta)(\theta^* - \theta - b(\theta)),\ \mu^n - p.p.,\ \theta \in [\theta_1, \theta_2].$$

Soit

$$\ln f_\theta(\mathbf{x}) - \ln f_{\theta_1}(\mathbf{x}) = \theta^* \int_{\theta_1}^{\theta} c(t)dt - \int_{\theta_1}^{\theta} c(t)a(t)dt$$

ce qui est (**) avec

$$h(\mathbf{x}) = f_{\theta_1}(\mathbf{x}),\ C(\theta) = \int_{\theta_1}^{\theta} c(t)dt,\ D(\theta) = -\int_{\theta_1}^{\theta} c(t)a(t)dt.$$

2) Si $\theta^* = \text{cst.}$, ou $b'(\theta) = -1$, il est clair que les deux membres de (*) sont nuls. Supposons (**) satisfaite. En dérivant la fonction log-vraisemblance $l(\mathbf{x}, \theta)$ par rapport à θ, on a :

$$\frac{d}{d\theta} \ln f_\theta(\mathbf{X}) = \theta^* C'(\theta) + D'(\theta).$$

D'où, puisque $\mathrm{E}_\theta[\frac{d}{d\theta} \ln f_\theta(\mathbf{X})] = 0 = A'(\theta)\,\mathrm{E}_\theta(\theta^*) + D'(\theta) = C'(\theta)(\theta + b(\theta)) + D'(\theta)$,

$$\frac{d}{d\theta} \ln f_\theta(\mathbf{X}) = C'(\theta)(\theta^* - \theta - b(\theta)),$$

ce qui prouve

$$\frac{f'_\theta(\mathbf{X})}{f_\theta(\mathbf{X})} = c(\theta)(\theta^* - \theta - b(\theta)), \text{ où } c(\theta) = C'(\theta)$$

et montre l'égalité dans (*). □

Corollaire 1.7 *Sous les conditions pour l'inégalité de Cramer-Rao, pour que la borne inférieure soit atteinte dans cette inégalité, il est nécessaire et suffisant que θ^* soit exhaustif et que la fonction $g(\theta^*, \theta)$ du théorème de factorisation soit de la forme*

$$g(\theta^*, \theta) = \exp\{\theta^* C(\theta) + D(\theta)\},$$

avec C et D dérivables.

Proposition 1.4 *Supposons qu'il existe l'estimateur de maximum de vraisemblance et qu'il est unique. Si les conditions pour l'inégalité de Cramer-Rao sont remplies et s'il existe un estimateur sans biais R-efficace, alors c'est l'estimateur de maximum de vraisemblance.*

Preuve. En effet, si θ^* est R-efficace,

$$\frac{d}{d\theta} \ln f_\theta(\mathbf{X}) = (\theta^* - \theta)\, c(\theta).$$

Puisque $b(\theta) = 0$, on a :

$$\mathrm{E}_\theta[(\theta^* - \theta)^2] = \mathrm{Var}_\theta(\theta^*)$$

et comme $\mathrm{E}_\theta[\frac{d}{d\theta} \ln f_\theta(\mathbf{X})^2] = nI(\theta)$,

$$c(\theta)^2 = \frac{nI(\theta)}{\mathrm{Var}_\theta(\theta^*)} = (nI(\theta))^2 > 0.$$

Ceci prouve que $c(\theta)$ garde un signe constant. Comme l'estimateur de maximum de vraisemblance existe et est unique on voit que $\frac{d}{d\theta} \ln f_\theta(\mathbf{X}) > 0$, pour $\theta < \theta^*$ et $\frac{d}{d\theta} \ln f_\theta(\mathbf{X}) < 0$, pour $\theta > \theta^*$. Ainsi $\ln f_\theta(\mathbf{X})$ atteint son maximum en θ^*. □

Passons au cas vectoriel, maintenant :

Théorème 1.25 *Soit $\theta^\star \in K_{\mathbf{b}}$ et supposons que les conditions pour l'inégalité de Cramer-Rao sont remplies. Supposons que la matrice de covariance*

$$\mathrm{Cov}_\theta(\theta^*) = \mathrm{E}_\theta\left[(\theta^* - \theta - \mathbf{b}(\theta))^\star(\theta^* - \theta - \mathbf{b}(\theta)\right]$$

existe et que son déterminant de est strictement positif (ou que le déterminant de $Id_k + B(\theta)$ est strictement positif) pour tout θ. Dans ce cas l'égalité est réalisée dans l'inégalité de Cramer-Rao si et seulement si la distribution de l'échantillon est exponentielle de type suivant :

$$f_\theta(\mathbf{x}) = \exp\left[< \theta^*, C(\theta) > + D(\theta)\right] h(\mathbf{x}),$$

où le vecteur $C(\theta) = (C_1(\theta), \ldots, C_k(\theta))$ satisfait

$$\left(\frac{\partial C_j(\theta)}{\partial \theta_\ell}\right)_{1 \leqslant j,\ell \leqslant k} = n\,(Id_k + B(\theta))^\star\, I(\theta).$$

Pour les estimateurs $\theta^\star$ sans biais, il est clair que

$$\mathrm{Cov}_\theta(\theta^*) \geqslant (nI(\theta))^{-1}$$

et l'égalité n'est possible que si le modèle est exponentiel avec

$$\left(\frac{\partial C_j(\theta)}{\partial \theta_\ell}\right)_{1 \leqslant j,\ell \leqslant k} = nI(\theta).$$

Remarque : Donc si on a réussit à trouver θ^* sans biais de matrice de covariance $(nI(\theta))^{-1}$, alors cet estimateur sera efficace. Notons que

$$\mathrm{E}_\theta(\theta^* - \theta)^\star(\theta^* - \theta) = \mathrm{Cov}_\theta(\theta^*) + \mathbf{b}(\theta)^\star\mathbf{b}(\theta).$$

Preuve. Posons

$$L'_j = L'_j(X,\theta) = \sum_{i=1}^n l'_j(X_i,\theta),\ \mathbf{L}' = \mathbf{L}'(\mathbf{X},\theta) = (L'_1,\ldots,L'_r).$$

On a $\mathrm{E}_\theta(\mathbf{L}') = 0$ et $\mathrm{E}_\theta(\theta^*)^\star\mathbf{L}' = Id_k + B(\theta)$, d'où

$$\mathrm{E}_\theta(\theta^* - \theta - b(\theta))^\star\mathbf{L}' = Id_k + B(\theta).$$

L'égalité en inégalité de Cramer-Rao n'est possible que si

$$(\theta^* - \theta - \mathbf{b}(\theta))^\star = (Id_k + B(\theta))(nI(\theta))^{-1}(\mathbf{L}')^\star,$$

ou encore

$$\mathbf{L}' = (\theta^* - \theta - \mathbf{b}(\theta))n\left[(Id_k + B(\theta))^{-1}\right]^\star I(\theta)$$

(le déterminant de $Id_k + B(\theta)$ est non nul par l'égalité dans l'inégalité et le fait que le déterminant de $\mathrm{Cov}_\theta(\theta^*)$ est non nul). Enfin on peut déduire que

$$\ln f_\theta(\mathbf{x}) = <\theta^*, C(\theta)> + D(\theta) + H(\mathbf{x}),$$

d'où

$$\mathbf{L}' = \theta^*\left(\frac{\partial C_j(\theta)}{\partial\theta_\ell}\right) + D'(\theta).$$

Mais $\mathrm{E}_\theta(\mathbf{L}') = 0$, donc $D'(\theta) = -\theta - \mathbf{b}(\theta)\left(\frac{\partial C_j(\theta)}{\partial\theta_\ell}\right)_{j,\ell}$. On multiplie à gauche par $(\theta^* - \theta - \mathbf{b}(\theta))^\star$ l'égalité

$$\mathbf{L}' = (\theta^* - \theta - \mathbf{b}(\theta))\left(\frac{\partial C_j(\theta)}{\partial\theta_\ell}\right)_{j,\ell},$$

donc pour que l'égalité soit réalisée il faut que

$$\left(\frac{\partial C_j(\theta)}{\partial\theta_\ell}\right)_{j,\ell} = n\left[(Id_k + B(\theta))^{-1}\right]^\star I(\theta).$$

□

La définition de la R-efficacité se modifie en conséquence et on a aussi

Proposition 1.5 *Supposons qu'il existe l'estimateur de maximum de vraisemblance et qu'il est unique. Supposons remplies les conditions pour que l'inégalité de Cramer-Rao soit vraie. Si θ^* est un estimateur sans biais R-efficace, alors c'est l'estimateur de maximum de vraisemblance.*

Preuve. Pour prouver qu'un estimateur sans biais R-efficace est le seul point de maximum de la vraisemblance, il suffit de s'assurer que $\mathbf{L}' = 0$ et que

$$< \nabla \ln f_\theta, u >=< \mathbf{L}', \theta^* - \theta > \quad \text{est négatif pour } \theta = \theta^* + u,\, u \neq 0.$$

Or l'égalité n'est possible dans l'inégalité de Cramer-Rao que si

$$\mathbf{L}' = (\theta^* - \theta) nI(\theta),$$

d'où on déduit les relations annoncées. La deuxième résulte du fait que $< \mathbf{L}', u >= -u\, nI(\theta)u^\star$, où $uI(\theta)u^\star$ est une forme quadratique positive définie. □

1.10 Estimation par intervalle (ou région) de confiance

Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un n-échantillon de loi $\mathrm{P}_\theta(\cdot)$, $\theta \in \Theta$. On continue de supposer satisfaites les hypothèses (A_0) et (A_μ). Dans ce paragraphe, $\alpha \in [0, 1]$ est fixé. Comme d'habitude on commence par le cas de la dimension 1 :

Définition 1.24 *Supposons $\Theta \subset \mathbb{R}$ et qu'on peut trouver deux fonction $a, b : \mathbb{R}^n \to \mathbb{R}$ telles que $a(\mathbf{X}) \leqslant b(\mathbf{X})$. Supposons de plus que l'intervalle aléatoire $[a(\mathbf{X}), b(\mathbf{X})]$ satisfait*

$$\mathrm{P}_\theta(a(\mathbf{X}) \leqslant \theta \leqslant b(\mathbf{X})) = 1 - \alpha, \forall \theta \in \Theta.$$

Alors, après la réalisation de l'expérience, $[a(\mathbf{x}^{obs}), b(\mathbf{x}^{obs})]$ est un **intervalle de confiance pour θ avec coefficient de sécurité** $1 - \alpha$.

L'interprétation du fait que $[a(\mathbf{X}), b(\mathbf{X})]$ est un intervalle de confiance est la suivante : on observe $\mathbf{x}^{obs}$ et on construit ainsi $[a(\mathbf{x}^{obs}), b(\mathbf{x}^{obs})]$ et on annonce que $\theta_0 \in [a(\mathbf{x}^{obs}), b(\mathbf{x}^{obs})]$. Si on répète cela N fois d'une manière indépendante, quand N est très grand, on devrait avoir raison au moins $(1 - \alpha)N$ fois parmi les N répétitions.

Une fois que l'expérience est réalisée l'intervalle $[a(\mathbf{x}^{obs}), b(\mathbf{x}^{obs})]$ est déterministe et alors, ou bien $\theta_0 \in [a(\mathbf{x}^{obs}), b(\mathbf{x}^{obs})]$, ou bien $\theta_0 \notin [a(\mathbf{x}^{obs}), b(\mathbf{x}^{obs})]$.

Pour des lois discrètes la probabilité de la définition précédente peut prendre seulement un nombre fini de valeurs. On cherche alors les fonction a et b telles que

$$\mathrm{P}_\theta(a(\mathbf{X}) \leqslant \theta \leqslant b(\mathbf{X})) \geqslant 1 - \alpha, \forall \theta \in \Theta$$

ou

$$\mathrm{P}_\theta(a(\mathbf{X}) \leqslant \theta \leqslant b(\mathbf{X})) \approx 1 - \alpha, \forall \theta \in \Theta.$$

Dans cette dernière situation on dit avoir construit un **intervalle de confiance approximatif** pour θ de coefficient de sécurité $1-\alpha$.

La méthode générale de construction d'intervalles de confiance est la suivante. Il faut trouver une variable aléatoire **fonction (ou variable aléatoire) pivotale** $T_n(\theta) = T_n(\mathbf{X}, \theta)$ qui dépend de θ et aussi de $\mathbf{X}$ mais seulement par l'intermédiaire d'une statistique suffisante pour θ dont on connaît la loi. On peut, ensuite, déduire les bornes de l'intervalle de confiance à partir de $T_n(\theta)$.

Il est possible qu'il y a plusieurs intervalles de confiance pour θ avec le même coefficient de sécurité $1-\alpha$. Il est clair qu'on sera intéressé par le plus court. Ce problème est souvent très compliqué.

Définition 1.25 *Une probabilité (ou une densité f de probabilité) sur $\mathbb{R}$ est* **unimodale autour d'un mode** *s'il existe $x^\star$* **un mode** *tel que la fonction de répartition de cette loi est convexe (f croissante) sur $]-\infty, x^\star]$ et concave (f décroissante) sur $[x^\star, +\infty[$.*

Exemples : $\mathcal{N}(m, \sigma^2)$ est unimodale autour de m ; la densité de loi de Cauchy $[2\pi(1+(x-m)^2]^{-1}$ autour de m ; la loi uniforme sur $[0,1]$ autour de tout $m \in [0,1]$.

Proposition 1.6 *Soit f une densité unimodale et soit $[a,b]$ satisfaisant :*

i) $\int_a^b f(x)dx = 1-\alpha$;

ii) $f(a) = f(b) > 0$;

iii) $a \leqslant x^\star \leqslant b$, *où $x^\star$ est un mode de f.*

Alors $[a,b]$ est l'intervalle le plus court parmi tous les intervalles satisfaisant i).

Preuve. Soit $[a', b']$ un autre intervalle avec $b'-a' < b-a$. On va montrer que cela implique forcement $\int_{a'}^{b'} f(x)dx < 1-\alpha$. On fait la preuve pour $a' \leqslant a$ (l'autre cas est identique). On peut avoir $b' \leqslant a$ ou $b' > a$.

Supposons que $a' \leqslant b' \leqslant a \leqslant x^\star$:

$$\begin{aligned}
\int_{a'}^{b'} f(x)dx &\leqslant f(b')(b'-a') \text{ car } x \leqslant b' \leqslant x^\star \Rightarrow f(x) \leqslant f(b') \\
&\leqslant f(a)(b'-a') \text{ car } b' \leqslant a \leqslant x^\star \Rightarrow f(b') \leqslant f(a) \\
&< f(a)(b-a) \text{ car } b'-a' < b-a \text{ et } f(a) > 0 \\
&\leqslant \int_a^b f(x)dx = 1-\alpha \text{ par ii), iii) et la propriété unimodale } f(x) \geqslant f(a), \forall a \leqslant x \leqslant b.
\end{aligned}$$

Supposons que $a' \leqslant a < b' < b$ (car si $b' > b$ alors $b'-a' \geqslant b-a$, absurde). On écrit

$$\int_{a'}^{b'} f(x)dx = \int_a^b f(x)dx + \left[\int_{a'}^{a} f(x)dx - \int_{b'}^{b} f(x)dx\right] = 1-\alpha + \left[\int_{a'}^{a} f(x)dx - \int_{b'}^{b} f(x)dx\right]$$

et va montrer que le crochet est négatif. Il est facile de voir que $\int_{a'}^{a} f(x)dx \leqslant f(a)(a-a')$, mais aussi que $\int_{b'}^{b} f(x)dx \geqslant f(b)(b-b')$. Montrons par exemple la deuxième inégalité quand $a' \leqslant a < b' \leqslant x^\star < b$. On a

$$\int_{b'}^{b} f(x)dx \geqslant \int_{b'}^{x^\star} f(x)dx + \int_{x^\star}^{b} f(x)dx \geqslant f(a)(x^\star - b') + f(b)(b - x^\star) = f(b)(b-b'),$$

par ii). Ainsi on trouve que

$$\begin{aligned} \int_{a'}^{a} f(x)dx - \int_{b'}^{b} f(x)dx &\leqslant f(a)(a-a') - f(b)(b-b') \\ &= f(a)\left[(a-a') - (b-b')\right] = f(a)\left[(b'-a') - (b-a)\right] < 0. \end{aligned}$$

□

Exemples :
1) Soit $\mathbf{X}$ un n-échantillon de loi gaussienne $\mathcal{N}(m, \sigma^2)$.
i) On suppose que σ est connu, donc le paramètre est m. Alors on prend $T_n(m) = \sqrt{n}(\bar{X} - m)/\sigma$ comme fonction pivotale. On a que $T_n(m) \sim \mathcal{N}(0,1)$ pour tout m. On trouve deux réels a, b tels que $a < b$ et

$$\mathrm{P}(a \leqslant G \leqslant b) = 1 - \alpha,\ G \sim \mathcal{N}(0,1),$$

d'où

$$\mathrm{P}_m(a \leqslant \frac{\sqrt{n}(\bar{X} - m)}{\sigma} \leqslant b) = 1 - \alpha,$$

donc

$$\mathrm{P}_m(\bar{X} - b\frac{\sigma}{\sqrt{n}} \leqslant m \leqslant \bar{X} - a\frac{\sigma}{\sqrt{n}}) = 1 - \alpha,$$

c'est-à-dire

$$\left[\bar{X} - b\frac{\sigma}{\sqrt{n}}, \bar{X} - a\frac{\sigma}{\sqrt{n}}\right]$$

est un intervalle de confiance de m de coefficient de sécurité $1-\alpha$. Sa longueur est $(b - a)\sigma/\sqrt{n}$. Il faut trouver a, b tels que $\Phi(b) - \Phi(a) = 1 - \alpha$ et tels que $b - a$ soit minimum. Il suffit de prendre $a = -c$, $b = c > 0$ où c est la quantile supérieure $\alpha/2$ de $\mathcal{N}(0,1)$ notée $g_{\alpha/2}$. Donc l'intervalle le plus court est

$$\left[\bar{X} - g_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + g_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right].$$

ii) On suppose que m est connu et que σ^2 est le paramètre. On prend la fonction pivotale

$$T_n(\sigma^2) = \frac{nS_1^2}{\sigma^2},\ \text{ avec }\ S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2.$$

La loi de cette fonction pivotale est $\chi^2(n)$ pour tout σ^2. On trouve deux réels $0 < a < b$ tels que :

$$P(a \leqslant Y \leqslant b) = 1 - \alpha,\ Y \sim \chi^2(n),$$

d'où

$$P_{\sigma^2}(a \leqslant \frac{nS_1^2}{\sigma^2} \leqslant b) = 1 - \alpha,$$

donc

$$P_{\sigma^2}(\frac{nS_1^2}{b} \leqslant \sigma^2 \leqslant \frac{nS_1^2}{a}) = 1 - \alpha,$$

c'est-à-dire

$$\left[\frac{nS_1^2}{b}, \frac{nS_1^2}{a}\right] = 1 - \alpha$$

est un intervalle de confiance de σ^2 de coefficient de sécurité $1 - \alpha$. Sa longueur est $(1/a - 1/b)nS_1^2$ et on veut la minimiser. On cherche $b = b(a)$ pour que cette longueur soit minimale, où $0 < a < b$ tels que

$$\int_a^b g_n(y)dy = 1 - \alpha,\ g_n \text{ densité de } \chi^2(n).$$

Le a pour lequel $l := (1/a - 1/b)nS_1^2$ est le plus petit est donné par $dl/da = 0$ qui est équivalent à $db/da = b^2/a^2$. De plus la condition précédente s'écrit aussi $G_n(b) - G_n(a) = 1 - \alpha$, où G_n est la fonction de répartition de $\chi^2(n)$. On dérive en a et on trouve

$$g_n(b)\frac{db}{da} - g_n(a) = 0 \text{ ou } \frac{db}{da} = \frac{g_n(a)}{g_n(b)}.$$

Ainsi la condition de longueur minimale s'écrit

$$a^2 g_n(a) = b^2 g_n(b),\ \int_a^b g_n(y)dy = 1 - \alpha.$$

D'ici on peut déduire avec des tables a, b.

2. Soit $\mathbf{X}$ un n-échantillon de loi $\mathcal{U}_{[0,\theta]}$ et on suppose $\theta > 0$. Alors on prend la fonction pivotale

$$T_n(\theta) = \frac{X_{(n)}}{\theta}$$

de densité $h_n(t) = nt^{n-1}\mathbb{1}_{[0,1]}(t)$. Un intervalle de confiance pour θ sera $[\frac{X_{(n)}}{b}, \frac{X_{(n)}}{a}]$ où $0 < a \leqslant b \leqslant 1$ sont tels que $b^n - a^n = 1 - \alpha$. On peut chercher le plus petit intervalle (exercice).

3. Soit $\mathbf{X}$ un n-échantillon de loi gaussienne $\mathcal{N}(m, \sigma^2)$.

Cette fois-ci σ^2 est inconnu (paramètre de nuisance) et on veut un intervalle pour m. On prend la fonction pivotale

$$T_n(m) = \frac{\sqrt{n}(\bar{X} - m)}{S_0} \sim \text{Student}(n-1)$$

où $S_0^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. L'intervalle de confiance est alors

$$\left[\bar{X} - b\frac{S_0}{\sqrt{n}}, \bar{X} - a\frac{S_0}{\sqrt{n}}\right],$$

où $a < b$ sont tels que

$$\mathrm{P}(a \leqslant T \leqslant b) = 1 - \alpha,\ T \sim \mathrm{Student}(n-1).$$

De la même façon, si m est inconnu (paramètre de nuisance) et on veut un intervalle pour σ^2, on prend

$$T_n(\sigma^2) = \frac{(n-1)S_0^2}{\sigma^2} \sim \chi^2(n-1)$$

et on trouve l'intervalle de confiance $[\frac{(n-1)S_0^2}{b}, \frac{(n-1)S_0^2}{a}]$, où

$$\mathrm{P}(a \leqslant Y \leqslant b) = 1 - \alpha,\ Y \sim \chi^2(n-1).$$

4. (intervalle de confiance approximatif)

Soit $\mathbf{X}$ un n-échantillon d'espérance m inconnue mais de variance σ^2 connue. On sait que, lorsque n est grand

$$\frac{\sqrt{n}(\bar{X} - m)}{\sigma} \approx \mathcal{N}(0,1)$$

D'ici on peut voir que, lorsque n est grand

$$\left[\bar{X} - g_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + g_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

est un intervalle de confiance approximatif pour m de coefficient de sécurité $1-\alpha$. Lorsque la variance σ^2 est connue, on peut voir que, lorsque n est grand

$$\frac{\sqrt{n}(\bar{X} - m)}{S} \approx \mathcal{N}(0,1),\ S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$$

et on peut encore déduire un intervalle de confiance approximatif pour m.

On peut utiliser la même idée pour construire des intervalles de confiance approximatifs pour l'estimateur du maximum de vraisemblance ou pour ceux construits par la méthode des moments. En effet, on peut utiliser la normalité asymptotique de ces estimateurs. Par exemple, on sait que l'estimateur du maximum de vraisemblance satisfait (sous certaines conditions)

$$\hat{\theta}_n^* \overset{\mathrm{P}_\theta-\text{p.s.}}{\longrightarrow} \theta \ \text{ et } \ \sqrt{n}(\hat{\theta}_n^* - \theta) \overset{\mathrm{P}_\theta-\text{loi}}{\longrightarrow} \mathcal{N}(0, \frac{1}{I(\theta)}).$$

Lorsque la fonction I est continue on peut estimer $I(\theta)$ par l'estimateur fortement convergent $I(\hat{\theta}_n^*)$. On en déduit, par exemple :

$$\mathrm{P}_\theta\left(|\hat{\theta}_n^* - \theta| \leqslant g_{\alpha/2}\sqrt{\frac{1}{nI(\hat{\theta}_n^*)}}\right) \approx 1 - \alpha,$$

d'où on déduit un intervalle de confiance approximatif pour θ.

Passons au cas vectoriel :

Définition 1.26 *Supposons $\Theta \subset \mathbb{R}^k$ et qu'on peut trouver des fonctions $a_j, b_j : \mathbb{R}^n \to \mathbb{R}$ telles que $a_j(\mathbf{X}) \leqslant b_j(\mathbf{X})$, $j = 1, \ldots, k$. On note*

$$I_j(\mathbf{X}) = [a_j(\mathbf{X}), b_j(\mathbf{X})],\ j = 1, \ldots, k,\ I(\mathbf{X}) = [a_1(\mathbf{X}), b_1(\mathbf{X})] \times \ldots \times [a_k(\mathbf{X}), b_k(\mathbf{X})].$$

Supposons de plus que

$$\mathrm{P}_\theta(\theta \in I(\mathbf{X})) = 1 - \alpha, \forall \theta \in \Theta.$$

Alors $I(\mathbf{x}^{obs})$ est une **région de confiance pour θ avec coefficient de sécurité** $1 - \alpha$.

Remarque : Si chaque $I_j(\mathbf{x}^{obs})$ est un intervalle de confiance pour θ_j de coefficient de sécurité $1-\alpha_j$, $j = 1, \ldots, k$ et si les couples de variables aléatoires $(a_1(\mathbf{X}), b_1(\mathbf{X})), \ldots, (a_k(\mathbf{X}), b_k(\mathbf{X}))$ sont indépendants, alors $I(\mathbf{x}^{obs})$ est de coefficient de sécurité $\prod_{j=1}^k (1 - \alpha_j)$. Si on n'a pas l'indépendance, on peut montrer que

$$\mathrm{P}_\theta(\theta \in I(\mathbf{X})) \geqslant 1 - \sum_{j=1}^{k} \mathrm{P}_\theta(\theta_j \notin I_j(\mathbf{X})) \geqslant 1 - \sum_{j=1}^{k} \alpha_j,$$

par l'inégalité de Bonferoni. Il suffit ensuite de choisir $\alpha_j = \frac{\alpha}{k}$, $j = 1, \ldots, k$ et trouver pour $I(\mathbf{x}^{obs})$ le coefficient de sécurité $1 - \alpha$.

Exemple

Soit $\mathbf{X}$ un n-échantillon de loi gaussienne $\mathcal{N}(m, \sigma^2)$. Ici le paramètre est $\theta = (m, \sigma^2)$. On sait que les variables aléatoires

$$\frac{\sqrt{n}(\bar{X} - m)}{\sigma} \sim \mathcal{N}(0, 1) \text{ et } \frac{(n-1)S_0^2}{\sigma^2} \sim \chi^2(n-1)$$

sont indépendantes. On trouve $c > 0$ et $0 < a < b$ tels que

$$\mathrm{P}(-c \leqslant G \leqslant c) = \sqrt{1 - \alpha},\ G \sim \mathcal{N}(0, 1) \text{ et } \mathrm{P}(a \leqslant Y \leqslant b) = \sqrt{1 - \alpha},\ Y \sim \chi^2(n-1).$$

Alors

$$\mathrm{P}_\theta\left(-c \leqslant \frac{\sqrt{n}(\bar{X} - m)}{\sigma} \leqslant c,\ a \leqslant \frac{(n-1)S_0^2}{\sigma^2} \leqslant b\right) = 1 - \alpha$$

ou

$$\mathrm{P}_\theta\left((m - \bar{X}) \leqslant \frac{c^2\sigma^2}{n},\ \frac{(n-1)S_0^2}{b} \leqslant \sigma^2 \leqslant \frac{(n-1)S_0^2}{a}\right) = 1 - \alpha.$$

D'ici on peut tracer la région de confiance pour θ.

1.11 Exercices

1.11.1 Lois de probabilité

1. (*lois normale, de Cauchy et log-normale*)
Soient $m \in \mathbb{R}$, $\sigma^2 > 0$ et on note $f_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]$, pour $x \in \mathbb{R}$.

a) Prouver que f_{m,σ^2} est une densité de probabilité. Il s'agit de la *loi normale* de paramètre $(m,\sigma^2) \in \mathbb{R} \times \mathbb{R}_+$, notée $\mathcal{N}(m,\sigma^2)$.

b) Calculer $\max_{x\in\mathbb{R}} f_{m,\sigma^2}(x)$ et $\lim_{|x|\to\infty} f_{m,\sigma^2}(x)$. Tracer sur le même dessin les courbes des densités $f_{0,\frac{1}{4}}$, $f_{0,1}$ et $f_{0,4}$.

c) Calculer $\mathrm{E}(e^{-rX})$, $r \in \mathbb{R}$, si X admet f_{m,σ^2} comme densité. En déduire $\mathrm{E}(X)$, $\mathrm{E}(X^2)$, $\mathrm{Var}(X)$.

d) La fonction de répartition de la *variable gaussienne standard* Z de densité $f_{0,1}$ est tabulée. Soit X comme au point précédent. Calculer à l'aide de ces tables $\mathrm{P}(1 < X < 8)$, lorsque $m = 5$ et $\sigma = 2$. Trouver $g > 0$, tel que $\mathrm{P}(|Z| > g) = 0,1$.

e) On continue de supposer $X \sim \mathcal{N}(m,\sigma^2)$, $\sigma > 0$ et $Z \sim \mathcal{N}(0,1)$. Exprimer en termes de Z la probabilité $p_k = \mathrm{P}(|X-m| \leqslant k\sigma)$ et donner les valeurs pour $k = 1,2,3,4$.

f) Soient Z_1 et Z_2 deux variables aléatoires gaussiennes standard et indépendantes. Montrer que le quotient Z_1/Z_2 est une variable aléatoire de *loi de Cauchy*.

g) On continue de supposer $X \sim \mathcal{N}(m,\sigma^2)$ et on pose $Y = e^X$. Trouver la densité de la variable Y (de *loi log-normale*). Tracer le graphe de cette densité de paramètres $(0,\frac{1}{2})$, $(\frac{1}{2},\frac{1}{2})$ et $(1,\frac{1}{2})$.

h) Montrer que si Y est de loi log-normale alors $\mathrm{E}(Y) = e^{m+\frac{\sigma^2}{2}}$, $\mathrm{E}(Y^2) = e^{2m+2\sigma^2}$ et $\mathrm{Var}(Y) = e^{2m}(e^{\sigma^2}-1)$.

i) Enfin, soit Y comme aux deux points précédents. Trouver la loi de $1/Y$.

2. (*loi gamma, de chi-deux, de Fisher et de Student*)
Soient $p,\lambda > 0$ et $f_{p,\lambda}(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} \mathbb{1}_{x\geqslant 0}$, où $\Gamma(p) := \int_0^\infty x^{p-1}e^{-x}dx$.

a) Prouver que $f_{p,\lambda}$ est une densité de probabilité. Il s'agit de la *loi gamma* de paramètre $(p,\lambda) \in \mathbb{R}_+ \times \mathbb{R}_+$, notée $\gamma_{p,\lambda}$.

b) Tracer les courbes des densités $f_{1,1}$, $f_{2,1}$, $f_{4,1}$, ainsi que $f_{2,2}$, $f_{2,1}$ et $f_{2,\frac{1}{2}}$.

c) Calculer $\mathrm{E}(e^{-rX})$, $r \geqslant 0$, si X admet $f_{p,\lambda}$ comme densité. En déduire $\mathrm{E}(X)$, $\mathrm{E}(X^2)$, $\mathrm{Var}(X)$.

d) Soient $X_1,\ldots,X_k$, variables aléatoires indépendantes de lois $\gamma_{p_1,\lambda},\ldots,\gamma_{p_k,\lambda}$, respectivement. Montrer que la loi de leur somme est $\gamma_{p_1+\ldots+p_k,\lambda}$.

e) Soient $Z_1,\ldots,Z_d$, variables aléatoires gaussiennes indépendantes. Montrer que la loi de $Y = Z_1^2 + \ldots + Z_d^2$ est $\gamma_{\frac{d}{2},\frac{1}{2}}$, qui est *la loi de chi-deux à d degrés de liberté* $\chi^2(d)$. Calculer $\mathrm{E}(Y)$, $\mathrm{E}(Y^2)$, $\mathrm{Var}(Y)$. Trouver, à l'aide des tables, $a,b > 0$ tels que $\mathrm{P}(Y < a) = \mathrm{P}(Y > b) = 0,05$, pour diverses valeurs de d.

f) étudier la régularité de la densité de la loi $\chi^2(d)$. Pour quelle argument cette densité est-elle maximum ? Trouver un équivalent de ce maximum quand $d \to \infty$.

g) Montrer que $(Y - d)/\sqrt{2d}$ converge en loi, lorsque $d \to \infty$, vers une loi gaussienne standard.

h) Soit $W \sim \mathcal{N}_d(m, K)$ une variable aléatoire gaussienne d-dimensionnelle, d'espérance m et de matrice de covariance K. Montrer que la variable $Q(W) = (W-m)^* K^{-1}(W-m)$ suit la loi $\chi^2(d)$.

i) Soient les variables aléatoires indépendantes $Z \sim \mathcal{N}(0,1)$ et $Y \sim \chi^2(d)$. On pose $T = Z/\sqrt{Y/d}$. La loi de la variable aléatoire T est une *loi de Student à d degrés de liberté*. Trouver la densité de T. Calculer $\mathrm{E}(T)$, $\mathrm{E}(T^2)$, $\mathrm{Var}(T)$. Calculer la limite de la densité de T_d, lorsque $d \to \infty$. Trouver, à l'aide des tables et pour divers valeurs de d, $t > 0$, tel que $\mathrm{P}(|T| > t) = 0,1$.

j) Soient X_1 et X_2 deux variables aléatoires indépendantes de lois $\chi^2(d_1)$ et $\chi^2(d_2)$. On pose $F = (X_1/d_1)/(X_2/d_2)$. La loi de la variable aléatoire F est une *loi de Fisher de paramètre* (d_1, d_2), notée $\mathcal{F}(d_1, d_2)$. Trouver la densité de F. Quelle est la loi de $1/F$? Calculer $\mathrm{E}(F)$, $\mathrm{E}(F^2)$, $\mathrm{Var}(F)$. Trouver, à l'aide des tables, $a, b > 0$ tels que $\mathrm{P}(F < a) = \mathrm{P}(F > b) = 0,05$ pour $d_1 = 5$ et $d_2 = 6$.

k) Soit T une variable aléatoire de loi de Student à d degrés de libertés. Quelle est la loi de dT^2 ?

l) La loi de Cauchy est-elle un cas particulier d'une des lois de cet exercice ?

3. (*lois gamma, beta, arcsinus et de Cauchy*)

a) Soient $p, q, \lambda > 0$. Désignons par X et Y deux variables aléatoires indépendantes de lois respectivement $\gamma(p, \lambda)$, $\gamma(q, \lambda)$. Trouver la densité du vecteur aléatoire $(U, V) := (X + Y, X/(X + Y))$.

b) Les variables aléatoires U et V sont-elles indépendantes ? On dit que V suit une *loi beta* notée $\mathrm{B}(p, q)$.

c) Soit $V \sim \mathrm{B}(1/2, 1/2)$. On dit que V est une variable aléatoire de *loi arcsinus*. Prouver que la variable aléatoire $1/V$ a la même loi que la variable aléatoire $1 + C^2$, où C suit une loi de Cauchy.

d) Soit $X \sim \gamma(1, 1)$ et $V \sim \mathrm{B}(1/2, 1/2)$ deux variables indépendantes (X est de loi exponentielle de paramètre 1). Montrer que : $2XV \sim G^2$, où G est une variable aléatoire gaussienne centrée réduite.

e) Soient $\alpha \in \mathbb{R}$, $\sigma^2 > 0$ et on note $f_{\alpha,\sigma^2}(x) = \frac{1}{\pi\sqrt{\sigma^2}} \frac{1}{1+\frac{(x-\alpha)^2}{\sigma^2}}$, $x \in \mathbb{R}$. Prouver que f_{α,σ^2} est une densité de probabilité. Il s'agit de la *loi de Cauchy* de paramètre $(\alpha, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$, notée $\mathcal{C}(\alpha, \sigma^2)$.

f) Si $C \sim \mathcal{C}(0, 1)$, montrer que $\mathrm{E}(e^{itC}) = e^{-|t|}$ et que $D = \alpha + \sigma C \sim \mathcal{C}(\alpha, \sigma^2)$. Montrer que $\mathrm{E}(e^{itD}) = e^{-it\alpha - \sigma|t|}$. L'espérance de D est-elle définie ?

f) Si $D \sim \mathcal{C}(\alpha, \sigma^2)$ et $E \sim \mathcal{C}(\beta, \tau^2)$ sont deux variables aléatoires indépendantes, alors $D + E \sim \mathcal{C}(\alpha + \beta, (\sigma + \tau)^2)$.

4. Soient $X_1, \ldots, X_n$, n variables aléatoires indépendantes de même loi gaussienne $\mathcal{N}(m, \sigma^2)$. On pose $\bar{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$ et $S_n^2 = \frac{1}{n} \sum_{j=1}^{n} (X_j - \bar{X}_n)^2$. Montrer que $\bar{X}_n \sim \mathcal{N}(m, \frac{\sigma^2}{n})$, $\frac{n}{\sigma^2} S_n^2 \sim \chi^2(n-1)$ et que ces variables aléatoires sont indépendantes.

1.11.2 Convergence des variables aléatoires réelles

1. On considère l'espace de probabilité $([0,1], \mathcal{B}_{[0,1]}, \lambda)$, où λ est la mesure de Lebesgue. Toutes les variables aléatoires seront définies sur cet espace de probabilité. Rappel :

$$\begin{array}{ccccc} \text{convergence p.s.} & \Longrightarrow & \text{convergence en prob.} & \Longrightarrow & \text{convergence en loi} \\ & & \Uparrow & & \\ & & \text{convergence dans } \mathrm{L}^2 & & \end{array}$$

a) Construire une suite de variables aléatoires $\{X_n : n \in \mathbb{N}\}$ telle que $X_n \to 0$ en probabilité et que X_n ne converge pas vers 0 presque sûrement, quand $n \to \infty$.

b) Construire une suite de variables aléatoires $\{X_n : n \in \mathbb{N}\}$ telle que $X_n \to 0$ dans L^1 mais pas dans L^2, quand $n \to \infty$.

c) Montrer que, de toute suite $\{X_n : n \in \mathbb{N}\}$ telle que $X_n \to 0$ en probabilité, on peut extraire une sous-suite $\{X_{n_k} : k \in \mathbb{N}\}$ telle que $X_{n_k} \to 0$ presque sûrement, quand $k \to \infty$.

d) Construire une suite de variables aléatoires $\{X_n : n \in \mathbb{N}\}$ telle que $X_n \to 0$ presque sûrement et que X_n ne converge pas vers 0 dans L^1.

e) Montrer que si X est presque sûrement constante c, alors la convergence en probabilité de la suite $\{X_n : n \in \mathbb{N}\}$ vers X équivaut à la convergence en loi de cette suite vers X.

2. Soient $\{X_n : n \in \mathbb{N}\}$ et $\{Y_n : n \in \mathbb{N}\}$ des suites de variables aléatoires et soit $g : \mathbb{R}^2 \to \mathbb{R}$ une fonction continue. Toutes les convergences sont pour $n \to \infty$.

a) Montrer que si $X_n \to X$ et $Y_n \to Y$ presque sûrement (resp. en probabilité), alors $g(X_n, Y_n) \to g(X, Y)$ presque sûrement (resp. en probabilité).

b) Si $X_n \to X$ et $Y_n \to Y$ en probabilité, alors $X_n + Y_n \to X + Y$, $aX_n + bY_n \to aX + bY$ (a, b constantes), $X_n Y_n \to XY$ et enfin, lorsque $\mathrm{P}(Y_n \neq 0) = \mathrm{P}(Y \neq 0) = 1$, $X_n / Y_n \to X/Y$ en probabilité.

c) Si $X_n \to X$ en loi et et $Y_n \to c$ en probabilité alors

- **i)** $X_n + Y_n \to X + c$ en loi ;
- **ii)** $X_n Y_n \to cX$ en loi ;
- **iii)** $X_n / Y_n \to X/c$ en loi, dès que $\mathrm{P}(Y_n \neq 0) = 1$ et $c \neq 0$.

3. Soit $\{X_n : n \in \mathbb{N}\}$ une suite i.i.d. de variables aléatoires d'espérance m et de variance σ^2. On pose $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ et $S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$.

a) Calculer les limites en probabilité et presque sûres de $\bar{X}$ et de S^2, quand $n \to \infty$.

b) Montrer que $\frac{n}{n-1}\frac{S^2}{\sigma^2} \to 1$, en probabilité et presque sûrement, quand $n \to \infty$.

c) Calculer la limite, quand $n \to \infty$, de la suite de fonctions

$$G_n(x) = \mathrm{P}\left[\frac{\sqrt{n}(\bar{X} - m)}{\sigma} \leqslant x\right], \, x \in \mathbb{R}$$

De quel type de convergence s'agit-il ?

d) Montrer que $\frac{\sqrt{n-1}}{\sqrt{S^2}}(\bar{X} - m) \to \mathcal{N}(0,1)$ et $\frac{\sqrt{n}}{\sqrt{S^2}}(\bar{X} - m) \to \mathcal{N}(0,1)$ en loi, quand $n \to \infty$.

Remarque : on peut montrer que la convergence du point **c)** *est uniforme en $x \in \mathbb{R}$ (résultat admis).*

4.

i) Soit $\phi : \mathbb{R} \to \mathbb{R}$ dérivable avec ϕ' continue en un point $a \in \mathbb{R}$. Soient $\{c_n : n \in \mathbb{N}\}$ une suite de réels, non-nuls, telle que $c_n \to \infty$, et $\{X_n : n \in \mathbb{N}\}$ une suite de variables aléatoires telle que $c_n(X_n - a) \to X$ en loi, quand $n \to \infty$. Montrer que $c_n(\phi(X_n) - \phi(a)) \to \phi'(a)X$ en loi, quand $n \to \infty$.

ii) Soit $\{X_n : n \in \mathbb{N}\}$ une suite i.i.d. de variables aléatoires d'espérance m et de variance σ^2. Soit $g : \mathbb{R} \to \mathbb{R}$ dérivable avec dérivée continue au point m. Montrer que $\sqrt{n}[g(\bar{X}) - g(m)] \to \mathcal{N}(0, [\sigma g'(m)]^2)$, en loi quand $n \to \infty$.

5.

a) Soit $\{X_n : n \in \mathbb{N}\}$ une suite de variables aléatoires gaussiennes centrées de variance σ_n^2, telles que $\sigma_n^2 \to 0$. Trouver la limite en loi de X_n.

b) Soit $\{X_n : n \in \mathbb{N}\}$ une suite i.i.d. de variables aléatoires de loi commune Q. Montrer que, presque sûrement, $\frac{1}{n}\sum_{i=1}^n \delta_{X_i} \to Q$ étroitement.

c) Montrer que $\frac{1}{n}\sum_{i=1}^n \delta_{k/n}$ converge étroitement vers la mesure de Lebesgue sur $[0,1]$, quand $n \to \infty$.

d) Énoncer le théorème de Lévy sur la convergence en loi.

e) Soit $\{X_n : n \in \mathbb{N}\}$ une suite de variables aléatoires gaussiennes de lois respectivement $\mathcal{N}(m_n, \sigma_n^2)$. On suppose que $m_n \to m$ et $\sigma_n^2 \to \sigma^2$. Montrer que $X_n \to \mathcal{N}(m, \sigma^2)$ en loi.

f) Soit $\{X_n : n \in \mathbb{N}\}$ une suite de variables aléatoires telle que $X_n \to Q$ en loi. On suppose que la loi Q ne charge pas les points : $Q(\{x\}) = 0$, pour tout $x \in \mathbb{R}$. Montrer que pour tous $a, b \in \bar{\mathbb{R}}$, $a < b$, $\mathrm{P}(X_n \in [a,b]) \to Q([a,b])$ et $\mathrm{P}(X_n \in]a,b[) \to Q(]a,b[)$.

6.

a) Soit $\{X_n : n \in \mathbb{N}\}$ une suite de variables aléatoires de lois binomiales, telles que $X_n \sim \mathcal{B}(n, p_n)$, avec $p_n \to 0$ et $np_n \to \lambda$ quand $n \to \infty$. Montrer que pour chaque $k \in \mathbb{N}$ fixé, $\mathrm{C}_n^k p_n^k (1-p_n)^{n-k} \to e^{-\lambda}\frac{\lambda^k}{k!}$.

b) Soit une suite de variables aléatoires $\{X_n : n \in \mathbb{N}^*\}$, telle que X_n prenne les valeurs $1 - 1/n$ et $1 + 1/n$ avec les probabilités $\frac{1}{2}$ et $\frac{1}{2}$ ("densité de masse de probabilité" $f_n(x) = \frac{1}{2}\mathbb{1}_{\{x=1-1/n\}} + \frac{1}{2}\mathbb{1}_{\{x=1+1/n\}}$). Montrer que X_n converge en loi, mais que la suite des f_n ne converge pas vers une "densité de masse de probabilité".

1.11.3 Statistiques d'ordre. Information de Kullback-Leibler.

1. Soient $X_1, \ldots, X_n$ n variables aléatoires réelles indépendantes de même loi admettant une densité f. Pour chaque $\omega \in \Omega$, on définit $\sigma(\omega)$ la permutation de $\{1, \ldots, n\}$ qui classe par ordre croissant $X_1(\omega), \ldots, X_n(\omega)$. Ainsi

$$X_{\sigma(\omega)(1)}(\omega) \leqslant X_{\sigma(\omega)(2)}(\omega) \leqslant \ldots \leqslant X_{\sigma(\omega)(n)}(\omega).$$

On note pour simplifier $X_{\sigma(\omega)(i)}(\omega) := X_{(i)}(\omega)$. Soit R_i le rang de X_i dans le classement précédent.

a) Montrer que la variable aléatoire $\omega \mapsto \sigma(\omega)$ est p.s. définie et donner sa loi de probabilité. Calculer, pour $1 \leqslant i \leqslant n$ et $1 \leqslant k \leqslant n$, $\mathrm{P}(\sigma(i) = k)$.

b) Trouver la densité du n-uplet $(X_{(1)}, \ldots, X_{(n)})$.

c) Trouver la loi de $(R_1, \ldots, R_n)$ et ensuite montrer que les deux n-uplets $(R_1, \ldots, R_n)$ et $(X_{(1)}, \ldots, X_{(n)})$ sont indépendants.

d) Calculer $\mathrm{E}(R_i)$, $\mathrm{Var}(R_i)$ et $\mathrm{Cov}(R_i, R_j)$.

e) Trouver la densité de $X_{(i)}$ $(1 \leqslant i \leqslant n)$. Écrire les cas particuliers $i = 1$ et $i = n$.

f) Trouver la densité du couple $X_{(i)}, X_{(j)}$, $i < j$. Écrire le cas particulier $(i, j) = (1, n)$.

g) Trouver la densité de $X_{(n)} - X_{(1)}$.

On suppose maintenant que la loi commune des $X_1, \ldots, X_n$ est la loi uniforme sur $[0, 1]$.

h) Montrer que, pour i fixé et lorsque $n \to \infty$, $nX_{(i)}$ converge en loi vers une loi gamma dont on précisera les paramètres.

i) Montrer que $-\ln X_{(i)}$ a la même loi que $\sum_{k=i}^n Y_k/k$ où les variables aléatoires Y_k sont indépendantes de même loi exponentielle de paramètre 1.

j) Montrer que les variables $Z_1, \ldots, Z_n$ définies par $Z_i = (X_{(i)}/X_{(i+1)})^i$ $(1 \leqslant i \leqslant n-1)$, $Z_n = (X_{(n)})^n$ sont indépendantes et de loi uniforme sur $[0, 1]$.

On suppose maintenant que la loi commune des $X_1, \ldots, X_n$ est la loi exponentielle de paramètre λ.

k) Montrer que les variables $Y_1 = nX_{(1)}$, $Y_i = (n+1-i)(X_{(i)} - X_{(i-1)})$ $(1 < i \leqslant n)$ sont indépendantes et que ce n-uplet $(Y_1, \ldots, Y_n)$ a la même loi que $(X_1, \ldots, X_n)$.

l) Si $X_1, \ldots, X_n$ sont les instants de panne de n machines identiques, mises en route simultanément, alors $X_{(i)}$ est donc l'instant de la i-ème panne observée. Montrer que les intervalles entre deux pannes observées sont indépendants.

On revient à la situation générale et on note $F(x) = \int_{-\infty}^{x} f(y)dy$.

m) Montrer que les variables aléatoires $U_i = [F(X_{(i)})/F(X_{(i+1)})]^i$ $(1 \leqslant i \leqslant n-1)$ et $U_n = [F(X_n)]^n$ sont indépendantes et de loi uniforme sur $[0,1]$. On pourra remarquer que les variables aléatoires $F(X_i)$ sont de loi uniforme sur $[0,1]$ et que les variables aléatoires $-\ln F(X_i)$ sont exponentielles de paramètre 1.

2. Trouver l'information de Kullback-Leibler $H(\mathrm{P}_2 \mid \mathrm{P}_1)$ dans les cas suivants :

a) $\mathrm{P}_i = \mathcal{E}(\lambda_i)$, $i = 1, 2$;

b) $\mathrm{P}_i = \mathcal{P}(\lambda_i)$, $i = 1, 2$;

c) $\mathrm{P}_i = \mathcal{B}(1, p_i)$, $i = 1, 2$;

d) $\mathrm{P}_i = \mathcal{N}(m_i, \sigma_i^2)$, $i = 1, 2$.

1.11.4 Estimateurs : construction, propriétés asymptotiques, R-efficacité

1. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi P_θ, $\theta \in \Theta$. Calculer l'espérance et la variance commune. Écrire l'expression de la vraisemblance. Trouver des estimateurs pour le paramètre θ par les deux méthodes (de substitution et du maximum de vraisemblance). Étudier leur R-efficacité. étudier leurs propriétés asymptotiques (convergences et R-efficacité).

a) $\Theta = [0,1]$ et $\mathrm{P}_\theta = \mathcal{B}(1, \theta)$ (loi de Bernoulli) ;

b) $\Theta = \mathbb{R}_+$ et $\mathrm{P}_\theta = \mathcal{P}(\theta)$ (loi de Poisson) ;

c) $\Theta = \mathbb{R}$ et $\mathrm{P}_\theta = \mathcal{N}(\theta, 1)$ (loi normale avec variance connue) ;

d) $\Theta = \mathbb{R}_+^*$ et $\mathrm{P}_\theta = \mathcal{N}(0, \theta^2)$ (loi normale avec espérance connue) ;

e) $\Theta = \mathbb{R} \times \mathbb{R}_+^*$ et $\mathrm{P}_\theta = \mathcal{N}(m, \sigma^2)$ (loi normale) ;

f) $\Theta = \mathbb{R}_+$ et $\mathrm{P}_\theta = \mathcal{E}(\theta)$ (loi exponentielle).

2. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi uniforme sur $[0, \theta]$, $\theta \in \Theta$. Donner un estimateur sans biais pour θ par la méthode des moments ainsi qu'un estimateur sans biais par la méthode de maximum de vraisemblance. étudier les comportements asymptotiques des variances de ces deux estimateurs. Qu'en pensez-vous ?

3. Pour estimer la proportion p, inconnue, d'un certain type de poisson dans un lac, on décide de pêcher jusqu'à obtenir n poissons de ce type. Soit N le nombre total de prises. Proposer un estimateur de p. Montrer que l'estimateur $(n-1)/(N-1)$ $(n \geqslant 2)$ est sans biais.

4. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{B}(1,p)$, $0 < p < 1$.

a) Trouver un estimateur de $p(1-p)$à l'aide de $\bar{X}$. Est-il efficace ?

b) On veut estimer p^a, où a est un entier. Soit le polynôme $Q_a(x) = x(x-1)\ldots(x-a+1)$. Montrer que $Q_a(\sum_{i=1}^n X_i)/Q_a(n)$ est un estimateur sans biais de p^a.

c) Trouver toutes les fonctions de p admettant un estimateur sans biais.

d) Donner un estimateur de $p/(1-p)$. Est-il possible de construire un estimateur sans biais de cette quantité ?

5. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{N}(\theta, \theta)$, $\theta > 0$.

a) Calculer l'information de Fisher du modèle.

b) Déterminer l'estimateur du maximum de vraisemblance de θ. Montrer qu'il est biaisé, mais asymptotiquement sans biais. Quelle est sa loi asymptotique ?

6. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $\gamma(1, 1/a)$,$a > 0$. Déterminer l'estimateur du maximum de vraisemblance de a. Cet estimateur est-il R-efficace ?

7. Soit la famille de probabilités $\{\mathrm{P}_\theta : \theta \in \Theta \subset \mathbb{R}\}$ satisfaisant les hypothèses (A_0), (A_μ) et (A_1) et soit $\lambda = q(\theta)$ un changement bijectif de paramètre $q : \Theta \to \Lambda \subset \mathbb{R}$. Montrer que l'information de Fisher associée à la famille de probabilités $\{\mathrm{Q}_\lambda : \lambda \in \Lambda\}$ satisfaisant aussi les hypothèses (A_0), (A_μ) et (A_1) est donnée par $J(\lambda) = \frac{1}{|q'(\theta)|^2} I(\theta)$.

8. Soit $(Y_1, \ldots, Y_n)$ un n-échantillon de loi log-normale de paramètres m et σ^2. On utilise une autre paire de paramètres, m et $\theta = \mathbb{E}(Y) = e^{m+\frac{\sigma^2}{2}}$. Déterminer $\hat{m}$, $\hat{\theta}$ les estimateurs du maximum de vraisemblance de ces paramètres. Montrer que $\hat{\theta}$ est asymptotiquement sans biais et étudier son R-efficacité.

9. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi uniforme sur $[\theta, \theta+1]$, $\theta \in \mathbb{R}$. Montrer que l'estimateur du maximum de vraisemblance de θ n'est pas unique. Trouver un estimateur de θ par la méthode des moments.

10. Soit $(X_1, \ldots, X_n)$ un n-échantillon de densité

$$f_{m,\sigma^2}(x) = \frac{1}{2}\left\{\frac{\exp\left(-x^2/2\right)}{\sqrt{2\pi}} + \frac{\exp\left(-(x-m)^2/2\sigma^2\right)}{\sqrt{2\pi\sigma^2}}\right\}$$

On veut montrer qu'il n'existe pas d'estimateur du maximum de vraisemblance pour $\theta = (m, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$.

a) écrire la vraisemblance du modèle, $(m, \sigma^2) \mapsto f_{m,\sigma^2}(\mathbf{x})$.

b) On fixe $k \in \{1, \ldots, n\}$. Calculer la limite de la vraisemblance au point $(m, \sigma^2) = (x_k, 0)$. On pourra calculer la limite de chaque facteur dans l'expression de la vraisemblance. Utiliser ce calcul pour conclure.

11. Soit $((X_1, Y_1), \ldots, (X_n, Y_n))$ un n-échantillon de loi $\mathcal{N}_2\left(\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}\right)$. Montrer que l'estimateur du maximum de vraisemblance pour le paramètre $\theta = (m_1, m_2, \sigma_1^2, \sigma_2^2, \rho)$ est $(\overline{X}, \overline{Y}, S_X^2, S_Y^2, \hat{\rho})$, où

$$\hat{\rho} = \left[\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})\right] / \left[\sqrt{S_X^2 S_Y^2}\right].$$

1.11.5 Statistiques exhaustives complètes, théorème de Lehmann-Scheffé, modèles exponentiels

1. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi P_θ, $\theta \in \Theta$. Indiquer des statistiques exhaustives pour le paramètre θ. S'agit-il des statistiques complètes ? Trouver ensuite des estimateurs efficaces pour θ.

a) $\Theta = [0, 1]$ et $\mathrm{P}_\theta = \mathcal{B}(1, \theta)$ (loi de Bernoulli) ;

b) $\Theta = \mathbb{R}_+$ et $\mathrm{P}_\theta = \mathcal{P}(\theta)$ (loi de Poisson) ;

c) $\Theta = \mathbb{R}$ et $\mathrm{P}_\theta = \mathcal{N}(\theta, 1)$ (loi normale avec variance connue) ;

d) $\Theta = \mathbb{R}_+^*$ et $\mathrm{P}_\theta = \mathcal{N}(0, \theta^2)$ (loi normale avec espérance connue) ;

e) $\Theta = \mathbb{R} \times \mathbb{R}_+^*$ et $\mathrm{P}_\theta = \mathcal{N}(m, \sigma^2)$ (loi normale) ;

f) $\Theta = \mathbb{R}_+$ et $\mathrm{P}_\theta = \mathcal{E}(\theta)$ (loi exponentielle).

2. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi uniforme sur $[\mu, \theta]$, $-\infty < \mu < \theta < \infty$.

a) Indiquer une statistique exhaustive pour le paramètre (μ, θ).

b) Supposons que μ est connu et que le seul paramètre est θ. Trouver une statistique exhaustive complète pour θ. Peut-on trouver l'estimateur efficace pour θ ? (On pourra faire $\mu = 0$.)

c) Supposons cette fois que $\theta = \mu + 1$, est que le seul paramètre est μ. Montrer que la statistique trouvée au premier point est exhaustive mais n'est pas complète.

3. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi géométrique de paramètre $p \in]0, 1[$. Trouver une statistique suffisante pour p. Est-elle complète ?

4. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{B}(1, 1/(1+\theta))$, $\theta > 0$. Montrer que $X_1+\ldots+X_n$ statistique exhaustive. Est-elle complète ?

5. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{N}(\theta, \theta)$, $\theta > 0$. Montrer que $(S, T) := (X_1 + \ldots + X_n, X_1^2 + \ldots + X_n^2)$ est une statistique suffisante pour θ. Calculer l'espérance de $S^2 - T$. Que peut-on conclure ?

6. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $\gamma(p, \lambda)$.

a) Montrer que $(X_1 \ldots X_n, X_1+\ldots+X_n)$ est une statistique suffisante pour le paramètre (p, λ).

b) Supposons ensuite que λ est connu. Donner une statistique suffisante pour p.

c) Montrer que $\overline{X}$ est une statistique exhaustive pour le paramètre $\theta > 0$, lorsque la densité de l'échantillon est $f_\theta(x) = \frac{1}{6\theta^4} x^3 e^{-x/\theta} \mathbb{1}_{]0,\infty[}(x)$.

7. Soit $((X_1, Y_1) \ldots, (X_n, Y_n))$ un n-échantillon de loi $\mathcal{N}_2\left(\begin{pmatrix} m \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho \\ \rho & \tau^2 \end{pmatrix}\right)$. Montrer que $(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i, \sum_{i=1}^n X_i^2, \sum_{i=1}^n Y_i^2, \sum_{i=1}^n X_i Y_i)$ est une statistique exhaustive pour $(m, \mu, \sigma^2, \tau^2, \rho)$.

8. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{B}(1, p)$. On prend comme paramètre la variance $\theta = p(1-p)$. Utiliser le théorème de Lehmann-Scheffé pour trouver l'estimateur efficace de θ.

9. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi $\mathcal{P}(\lambda)$. On fixe $r \in \mathbb{N}$ et soit $\theta_r = e^{-\lambda} \lambda^r / (r!)$.

a) Soit $r = 0$ et on note $S = X_1 + \ldots + X_n$ et $N_0 = \sum_{i=0}^n \mathbb{1}_{\{X_i = 0\}}$.

i) On pose $T = \mathbb{E}(\frac{1}{n} N_0 \mid S)$. Montrer que $T = \mathbb{E}(\mathbb{1}_{\{X_1=0\}} \mid S)$.

ii) Montrer que $\mathbb{E}(\mathbb{1}_{\{X_1=0\}} \mid S = s) = \mathbb{P}(X_1 = 0 \mid S = s)$, $s \in \mathbb{N}$.

iii) En déduire que $\mathbb{E}(\mathbb{1}_{\{X_1=0\}} \mid S = s) = \left(\frac{n-1}{n}\right)^s$, $s \in \mathbb{N}$ et que $T = \left(1 - \dfrac{1}{n}\right)^{n\overline{X}}$.

iv) Montrer que T est un estimateur sans biais et efficace de θ_0.

b) Reprendre un raisonnement identique pour trouver un estimateur sans biais efficace de θ_r, pour $r \in \mathbb{N}^*$.

10. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi gaussienne $\mathcal{N}(m, \sigma^2)$. Trouver l'estimateur efficace de $p := \Phi(-\mu/\sigma)$, où Φ est la fonction de répartition d'une variable gaussienne standard (la fonction de Laplace). On pourra considérer d'abord le cas où σ^2 est connu.

11. Soient $(T_1, \ldots, T_r)$, r estimateurs sans biais d'un paramètre θ et on note $\mathrm{cov}(T_i, T_j) = \sigma_{ij}$, $i, j \in \{1, \ldots, r\}$.

a) Parmi les combinaisons linéaires des T_i, trouver l'estimateur sans biais de variance minimale.
b) Si $\sigma_{ij} = 0$ pour $i \neq j$, que vaut cette variance minimale ?
c) On dispose de r échantillons indépendants de tailles n_i, $i = 1, \ldots, r$, de lois $\mathcal{N}(m_i, \sigma^2)$. Proposer un estimateur sans biais de σ^2. Est-il efficace ?

12. Indiquer lesquels des modèles suivants sont des modèles exponentiels :

$\{\mathcal{B}(1,\theta) : \theta \in [0,1])\}$;$\{\mathcal{P}(\theta) : \theta > 0)\}$; $\{\mathcal{G}(\theta) : \theta \in [0,1])\}$; $\{\mathcal{E}(\theta) : \theta > 0)\}$;

$\{\gamma(p,\lambda) : \theta = (p,\lambda) \in \mathbb{R}_+^* \times \mathbb{R}_+^*\}$; $\{\mathcal{N}(m,\sigma^2) : \theta = (m,\sigma^2) \in)\mathbb{R} \times \mathbb{R}_+^*\}$;

$$\left\{\mathcal{N}_2\left(\begin{pmatrix} m \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho \\ \rho & \tau^2 \end{pmatrix}\right) : (m,\mu,\sigma^2,\tau^2,\rho) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+^* \times \mathbb{R}_+^* \times \mathbb{R}\right\}.$$

1.11.6 Intervalles de confiance

1. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi P_θ, $\theta \in \Theta$. Trouver des fonctions pivotales et ensuite déduire des intervalles (ou régions) de confiance de coefficient de sécurité $1 - \alpha \in [0,1]$ pour le paramètre θ.
a) $\Theta = \mathbb{R}$ et $\mathrm{P}_\theta = \mathcal{N}(\theta, 1)$ (loi normale avec variance connue) ;
b) $\Theta = \mathbb{R}_+^*$ et $\mathrm{P}_\theta = \mathcal{N}(0, \theta^2)$ (loi normale avec espérance connue) ;
c) $\Theta = \mathbb{R} \times \mathbb{R}_+^*$ et $\mathrm{P}_\theta = \mathcal{N}(m, \sigma^2)$ (loi normale) ;
d) $\Theta = \mathbb{R}_+$ et $\mathrm{P}_\theta = \mathcal{E}(\theta)$ (loi exponentielle).

2. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi uniforme sur $[0, \theta]$, $\theta \in \Theta$. Trouver une fonction pivotale et ensuite indiquer un intervalle de confiance pour θ de coefficient de sécurité $1 - \alpha \in [0,1]$.

3. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi P_θ, $\theta \in \Theta$. Trouver des intervalles de confiance asymptotiques pour θ de coefficient de sécurité $1 - \alpha \in [0,1]$.
a) $\Theta = [0,1]$ et $\mathrm{P}_\theta = \mathcal{B}(1, \theta)$ (loi de Bernoulli) ;
b) $\Theta = \mathbb{R}_+$ et $\mathrm{P}_\theta = \mathcal{P}(\theta)$ (loi de Poisson).

4. Soit $((X_1, Y_1), \ldots, (X_n, Y_n))$ un n-échantillon de loi $\mathcal{N}_2\left(\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right)$.
a) On suppose que $\sigma_1^2 = \sigma_2^2$ (inconnues) et on veut trouver un intervalles de confiance de coefficient de sécurité $1 - \alpha \in [0,1]$ pour le paramètre $\theta = m_1 - m_2$. On pourra d'abord indiquer la loi de la fonction pivotale

$$T_n(\theta) = \left[(\overline{X} - \overline{Y}) - \theta\right] / \sqrt{S_X^2 + S_Y^2}.$$

b) On ne suppose plus l'égalité des variances et on veut trouver un intervalles de confiance de coefficient de sécurité $1-\alpha \in [0,1]$ pour le paramètre $\tau = \sigma_1^2/\sigma_2^2$. On pourra d'abord indiquer la loi de la fonction pivotale

$$T_n(\tau) = \tau \, \frac{S_{0,X}^2}{S_{0,Y}^2}.$$

5. Soit $(X_1,\dots,X_n)$ un n-échantillon de loi $\mathcal{N}(\theta,1)$ (voir aussi l'exercice **1.a**). Montrer que l'intervalle aléatoire $[\overline{X}-1, \overline{X}+1]$ contient θ avec une probabilité qu'on calculera. Cette probabilité dépend-t-elle de n ? Trouver ensuite le plus petit n pour garantir un intervalle de confiance pour θ de longueur au plus $1/4$ de coefficient de sécurité 0,95. Si on suppose que la loi est $\mathcal{N}(\theta,\sigma^2)$, avec σ^2 inconnu, trouver le plus petit n pour garantir un intervalle de confiance pour θ de longueur au plus $\sigma/4$ de coefficient de sécurité 0,90.

6. Soit $(X_1,\dots,X_n)$ un n-échantillon de loi uniforme sur $[0,\theta]$, $\theta \in \Theta$ et soit $X_n = \max\{X_1,\dots,X_n\}$ (voir aussi l'exercice **2**). On veut construire deux types d'intervalles aléatoires contenant θ : $[aX_n, bX_n]$, $1 \leqslant a < b$ et $[X_n+c, X_n+d]$, $0 \leqslant c < d$. Calculer $\mathrm{P}_\theta(\theta \in [aX_n, bX_n])$ ainsi que $\mathrm{P}_\theta(\theta \in [X_n+c, X_n+d])$. Lequel préférer ?

7. Soit $(X_1,\dots,X_n)$ un n-échantillon de loi $\mathcal{N}(m,\sigma^2)$, où les deux paramètres sont inconnus. Trouver un intervalle de confiance de coefficient de sécurité $1-\alpha$, pour σ^2 en termes de deux constantes $a,b>0$ (voir aussi l'exercice **1 b**). Trouver les conditions qui doivent satisfaire a et b pour que l'intervalle soit le plus court possible. Donner ensuite les valeurs de a et b lorsque $\alpha = 0,1$ et $n=3$. Comparer l'intervalle obtenu avec celui qu'on trouve en utilisant les le réel $y_{\alpha/2}$ tel que $\mathrm{P}(Y > y_{\alpha/2}) = \alpha/2$, $Y \sim \chi^2(n-1)$.

8. Soit $(X_1,\dots,X_n)$ un n-échantillon de loi $\mathcal{B}(1,\theta)$. Reprendre la partie a) de l'exercice **3** en utilisant la quantité $\theta(1-\theta)$ a la place de $\overline{X}(1-\overline{X})$ pour rechercher l'intervalle de confiance.

9. Soient $a,b : \mathbb{R}^n \to \mathbb{R}$ telles que $a(\mathbf{x}) \leqslant b(\mathbf{x})$ pour tout $\mathbf{x}$, et $\mathrm{P}_\theta(a(\mathbf{X}) \leqslant \theta) = 1-\alpha_1$ et $\mathrm{P}_\theta(b(\mathbf{X}) \geqslant \theta) = 1-\alpha_2$. Calculer $\mathrm{P}_\theta(a(\mathbf{X}) \leqslant \theta \leqslant b(\mathbf{X}))$.

10. Soit $(X_1,\dots,X_n)$ un n-échantillon de loi $\mathcal{N}(m,\sigma^2)$, où les deux paramètres sont inconnus (voir aussi l'exercice **1.c**). On veut trouver des régions de confiance pour $\theta = (m,\sigma^2)$ en utilisant

a) les deux intervalles $[\overline{X}-k\frac{S_0}{\sqrt{n}}, \overline{X}+k\frac{S_0}{\sqrt{n}}]$ et $[\frac{(n-1)S_0^2}{b}, \frac{(n-1)S_0^2}{a}]$; choisir k,a,b tels que le région de confiance soit de coefficient de sécurité $1-\alpha$.

b) les deux intervalles $[\overline{X}-k\sqrt{\frac{\sigma^2}{n}}, \overline{X}+k\sqrt{\frac{\sigma^2}{n}}]$ et $[\frac{(n-1)S_0^2}{b}, \frac{(n-1)S_0^2}{a}]$; choisir k,a,b tels que le région de confiance soit de coefficient de sécurité $1-\alpha$.

c) Comparer les deux régions de confiance.

11.

a) Un échantillon aléatoire de 10 prélévements indépendants est tiré d'une population distribuée suivant une loi normale. Les valeurs sont :

$$1,19;\ 1,08;\ 1,18;\ 1,13;\ 1,16;\ 1,20;\ 1,15;\ 1,13;\ 1,10;\ 1,14.$$

Calculer l'intervalle de confiance au niveau 5% de l'espérance et de la variance de la population.

b) Une expérimentation sur les globulines a donné les résultats suivants :

$$\text{centres des classes : } 6;\ 8;\ 10;\ 12;\ 14;\ 16;\ 18;\ 20;\ 22;\ 24;\ 26$$

$$\text{effectifs des classes : } 2;\ 6;\ 13;\ 17;\ 17;\ 38;\ 10;\ 17;\ 6;\ 5;\ 2.$$

Calculer l'intervalle de confiance au niveau 5% de l'espérance et de la variance de la population (distribuée suivant une loi normale).

12. Sur 4000 naissances une enquête relève 2065 garçons. Donner un intervalle de confiance pour la proportion de garçons, au niveau 5% et 1%.

13. On a fait un sondage auprès de 900 personnes sur une possible modification de la Constitution. Les opinions favorables représentaient 40,1% des réponses.

a) Déterminer un intervalle de confiance asymptotique de niveau de confiance 95% pour la probabilité d'une réponse favorable.

b) à la suite d'une intense campagne d'explication en faveur de cette modification on va de nouveau faire un sondage, mais avec pour objectif l'évaluation de l'efficacité de la campagne et non l'estimation de la proportion de personnes favorables. La campagne aura été *vraiment efficace* si l'opinion favorable est devenue majoritaire. Combien de personnes devra-t-on interroger si on veut différencier avec des risques de 5% les situations : "la campagne n'a eu aucune efficacité" contre "la campagne a été *vraiment efficace*" ?

Chapitre 2

Théorie des tests d'hypothèse

2.1 Introduction et définitions générales

On considère un modèle paramétrique $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\}$ et un n-échantillon d'observations $\mathbf{X} = (X_1, \ldots, X_n)$ de loi dans $\mathcal{P}$.

Définition 2.1 *Une affirmation concernant le paramètre θ de type $\theta \in \Theta_0 \subset \Theta$ est une* **hypothèse** *(statistique) à propos de θ, notée de manière usuelle par* (H) *(ou* (H_0)*). L'affirmation que $\theta \in \Theta_1 \subset \Theta_0^c$ est aussi une hypothèse statistique nommée* **alternative** *de* (H) *(ou* (H_0)*) et est notée* (A) *(ou* (H_1)*). Si Θ_0 contient un seul point alors* (H) *est dite hypothèse* **simple** *; sinon on l'appelle* **composite**. *Même chose est vraie pour les alternatives.*

Définition 2.2 *Le problème de* **test** *est de décider laquelle des deux hypothèses est la vraie, à partir de l'observation de l'échantillon. Un test est une règle de décision qui précise pour quelles valeurs de l'échantillon on prend la décision d'accepter* (H) *comme vraie et pour quelles valeurs de l'échantillon la décision est de rejeter* (H) *et d'accepter* (A). *Ces dernières valeurs constituent la* **région de rejet** *(ou* **région critique***) et c'est un borélien $\mathcal{R}$ de $\mathbb{R}^n$. Son complémentaire est* **région d'acceptation**. *Préciser $\mathcal{R}$ équivaut à préciser l'indicatrice de $\mathcal{R}$, donc la fonction :*

$$\phi(\mathbf{x}) = \begin{cases} 1 & si\ \ \mathbf{x} \in \mathcal{R} \\ 0 & si\ \ \mathbf{x} \in \mathcal{R}^c. \end{cases}$$

Il s'agit de la **fonction de test** *(ou simplement du* **test***) ϕ. C'est un* **test pur**. *Un* **test randomisé** *(ou une* **fonction de test***) pour tester* (H) *contre* (A) *est une fonction borélienne $\phi : \mathbb{R}^n \to [0, 1]$ ayant l'interprétation suivante : si $\phi(\mathbf{x}) = p$, alors une pièce qui montre pile avec probabilité p est lancée et* (H) *est rejetée (acceptée) si pile (respectivement face) apparaît. Le cas particulier où $p = 0$ ou $p = 1$ est le* **test pur**.

Une méthode générale de construction des tests (ou des régions de rejet) sera décrite au §**2.4**. Elle basée sur une idée intuitive lorsque les deux hypothèses sont simples et généralisée pour toute situation. Une fois construits il faut savoir comparer les tests. Pour cela on a besoin de plusieurs notions.

Dans un test d'hypothèse on peut commettre deux types d'erreur : rejeter (H) (à tort) quand elle est en fait vraie ; accepter (H) quand elle est en fait fausse. Les conséquences ne sont pas identiques (plus ou moins graves).

Définition 2.3 *La* **fonction puissance** *associée à un test ϕ est l'application $\beta_\phi : \Theta \to [0,1]$, donnée par :*

$$\theta \mapsto \mathrm{P}_\theta(\text{ rejeter (H) }) = \mathrm{P}_\theta(\mathbf{X} \in \mathcal{R}) = \mathrm{E}_\theta\left[\phi(\mathbf{X})\right].$$

La probabilité

$$\mathrm{P}_\theta(\text{ rejeter (H) }) = \mathrm{P}_\theta(\mathbf{X} \in \mathcal{R}) = \mathrm{E}_\theta\left[\phi(\mathbf{X})\right], \forall \theta \in \Theta_0$$

est le **risque de 1ère espèce**. *Le* **risque de 2ième espèce** *est*

$$\mathrm{P}_\theta(\text{ accepter (H) }) = \mathrm{P}_\theta(\mathbf{X} \in \mathcal{R}^c) = 1 - \mathrm{P}_\theta(\mathbf{X} \in \mathcal{R}), \forall \theta \in \Theta_1.$$

Ainsi la fonction puissance donne le risque de 1ère espèce sur Θ_0 et donne 1 moins le risque de 2ième espèce sur Θ_1.

L'idéal est que β_ϕ soit nulle sur Θ_0 et 1 sur Θ_1. Un bon test a la fonction puissance proche de 0 pour la plupart des $\theta \in \Theta_0$ et proche de 1 pour la plupart des $\theta \in \Theta_1$. Cela n'est pas possible lorsque l'échantillon est fixé.

Définition 2.4 *Soit $\alpha \in]0,1[$ et soit $\phi : \mathbb{R}^n \to [0,1]$ un test de* (H) *$\theta \in \Theta_0$ contre* (A) *$\theta \in \Theta_1$. Le test est dit* **de niveau** α **pour** Θ_0 *si $\sup_{\theta \in \Theta_0} \beta_\phi(\theta) = \alpha$. Le test est dit* **de seuil** α **pour** Θ_0 *si $\sup_{\theta \in \Theta_0} \beta_\phi(\theta) \leqslant \alpha$.*

Définition 2.5 *Un test ϕ de niveau (ou seuil) α pour Θ_0 qui maximise la puissance parmi tous les tests de niveau (ou seuil) α est dit* **uniformément plus puissant (u.p.p.) contre** Θ_1 *:*

$$\sup_{\theta_0 \in \Theta_0} \beta_\phi(\theta) = \alpha \;\; \text{ou} \;\; \leqslant \alpha,$$

$$\beta_\phi(\theta) \geqslant \beta_{\phi*}(\theta),\ \forall \theta \in \Theta_1,\ \forall \phi^* \text{ de niveau ou de seuil } \alpha \text{ pour } \Theta_0.$$

Si Θ_1 a un seul point (alternative simple) le test est dit simplement **plus puissant (p.p.) pour** Θ_1

Il n'y a pas de raison à priori pour qu'il existe un test u.p.p.

Définition 2.6 *Un test ϕ pour tester*

$$(\mathrm{H}) : \theta \in \Theta_0 \;\; \text{contre} \;\; (\mathrm{A}) : \theta \in \Theta_1$$

est dit **sans biais** *si*

$$\beta_\phi(\theta'') \geqslant \beta_\phi(\theta'), \forall \theta' \in \Theta_0 \;\; \text{et} \;\; \forall \theta'' \in \Theta_1.$$

Ainsi le risque de 1ère espèce est au plus α et la puissance est au moins α. Un test est dit **uniformément plus puissant sans biais (u.p.p.s.b.) contre** Θ_1 *s'il est u.p.p. contre Θ_1 dans la classe des tests sans biais contre Θ_1.*

Remarque. Un test u.p.p. contre Θ_1 de seuil α pour Θ_0 est toujours un test u.p.p.s.b. contre Θ_1 de seuil α pour Θ_0. En effet il est sans biais contre Θ_1 car il est au moins tout aussi puissant que le test identiquement égal à α. Enfin il est u.p.p.s.b. contre Θ_1 car il est u.p.p. contre Θ_1 dans une classe contenant la classe des tests sans biais contre Θ_1.

2.2 Comparaison des tests

2.2.1 Tester une hypothèse simple contre une alternative simple

On commence par une situation simple où $\Theta = \{\theta_0, \theta_1\}$. En fait il peut y avoir plusieurs points mais on se focalise seulement sur ces deux points. On suppose que l'échantillon $\mathbf{X} = (X_1, \ldots, X_n)$ est de densité $f_\theta(\cdot)$ et notera plus simplement $f_i := f_{\theta_i}$, $i = 0, 1$. Le problème est de tester :

$$(\mathrm{H}) : \theta \in \Theta_0 = \{\theta_0\} \text{ contre } (\mathrm{A}) : \theta \in \Theta_0^c = \{\theta_1\}$$

au niveau α. Autrement dit, on veut tester si la densité de l'échantillon est f_0 ou f_1. Dans tout ce qui suit on notera la densité de l'échantillon

$$f_\theta(\mathbf{x}) = \prod_{j=1}^{n} f_\theta(x_j),\ \mathbf{x} = (x_1, \ldots, x_n).$$

Théorème 2.1 *(lemme fondamental de Neyman-Pearson)*
Soit $\mathbf{X}$ un n-échantillon de densité $f_\theta(\cdot)$, $\theta \in \Theta = \{\theta_0, \theta_1\}$. On veut tester

$$(\mathrm{H}) : \theta \in \Theta_0 = \{\theta_0\} \text{ contre } (\mathrm{A}) : \theta \in \Theta_0^c = \{\theta_1\}$$

au seuil $\alpha \in]0, 1[$ pour $\{\theta_0\}$. On définit le test

$$\phi(\mathbf{x}) = \begin{cases} 1, & si\ f_1(\mathbf{x}) > c\, f_0(\mathbf{x}) \\ \gamma, & si\ f_1(\mathbf{x}) = c\, f_0(\mathbf{x}) \\ 0, & sinon, \end{cases} \tag{1}$$

où les constantes $\gamma \in [0, 1]$ et $c > 0$ sont déterminées tels que

$$\mathrm{E}_{\theta_0}[\phi(\mathbf{X})] = \alpha. \tag{2}$$

Alors pour tester (H) *contre* (A) *tout test ϕ défini par (1)+(2) est le p.p. contre $\{\theta_1\}$ dans la classe des tests de seuil α pour $\{\theta_0\}$.*

Preuve. Soient

$$E = \{\mathbf{x} \in \mathbb{R}^n : f_0(\mathbf{x}) > 0\} \text{ et } D^c = \mathbf{X}^{-1}(E^c) = \mathbf{X}^{-1}(\{\mathbf{x} \in \mathbb{R}^n : f_0(\mathbf{x}) = 0\}).$$

On a alors

$$\mathrm{P}_{\theta_0}(D^c) = \mathrm{P}_{\theta_0}(\mathbf{X} \in E^c) = \int_{E^c} f_0(\mathbf{x})\mu(d\mathbf{x}) = 0,$$

donc dans les calculs de probabilité P_{θ_0} on peut redéfinir et modifier les variables aléatoires sur D^c. Alors on a, en particulier,

$$\mathrm{E}_{\theta_0}[\phi(\mathbf{X})] = \mathrm{P}_{\theta_0}[f_1(\mathbf{X}) > c\, f_0(\mathbf{X})] + \gamma\, \mathrm{P}_{\theta_0}[f_1(\mathbf{X}) = c\, f_0(\mathbf{X})]$$

$$= \mathrm{P}_{\theta_0}[\{f_1(\mathbf{X}) > c\, f_0(\mathbf{X})\} \cap D] + \gamma\, \mathrm{P}_{\theta_0}[\{f_1(\mathbf{X}) = c\, f_0(\mathbf{X})\} \cap D]$$

$$= \mathrm{P}_{\theta_0}\left[\left\{\frac{f_1(\mathbf{X})}{f_0(\mathbf{X})} > c\right\} \cap D\right] + \gamma\, \mathrm{P}_{\theta_0}\left[\left\{\frac{f_1(\mathbf{X})}{f_0(\mathbf{X})} > c\right\} \cap D\right]$$

$$= \mathrm{P}_{\theta_0}[\{T(\mathbf{X}) > c\} \cap D] + \gamma\, \mathrm{P}_{\theta_0}[\{T(\mathbf{X}) = c\} \cap D] = \mathrm{P}_{\theta_0}(T(\mathbf{X}) > c) + \gamma\, \mathrm{P}_{\theta_0}(T(\mathbf{X}) = c),$$

où $T(\mathbf{X}) := \frac{f_1(\mathbf{X})}{f_0(\mathbf{X})}$ sur D et $T(\mathbf{X})$ arbitraire sur D^c. Soit $G(t) = 1 - F(t) = \mathrm{P}_{\theta_0}(T(\mathbf{X}) > c)$. Alors $\mathrm{P}_{\theta_0}(T(\mathbf{X}) = c) = G(c-) - G(c)$, G est décroissante, continue à droite avec $G(-\infty) = 1$, $G(\infty) = 0$. Comme $T(\mathbf{X})$ est un rapport de densités de probabilités, il s'agit d'une variable positive $\mathrm{P}_{\theta_0}(T(\mathbf{X}) > 0) = 1$, donc $G(t) = 1$ si $t < 0$.

Pour tout $\alpha \in]0,1[$, il existe $c_0 \geqslant 0$ telle que $G(c_0) \leqslant \alpha \leqslant G(c_0-)$. En effet, ou bien $G(c_0) = G(c_0-)$, c'est-à-dire c_0 est un point de continuité de G, et alors $\alpha = G(c_0)$ et si on prend dans (1) $c = c_0$:

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } f_1(\mathbf{x}) > c\, f_0(\mathbf{x}) \\ 0, & \text{sinon} \end{cases}$$

(c'est-à-dire $\gamma = 0$). Le test est de niveau α pour $\{\theta_0\}$ car $\mathrm{E}_{\theta_0}[\phi(\mathbf{X})] = G(c_0) + 0 = \alpha$. Ou bien on a c_0 est un point de discontinuité de G et on prend $c = c_0$ et on pose

$$\gamma = \frac{\alpha - G(c_0)}{G(c_0-) - G(c_0)} \in [0,1].$$

A nouveau le test est de niveau α pour $\{\theta_0\}$

$$\mathrm{E}_{\theta_0}[\phi(\mathbf{X})] = G(c_0) + \frac{\alpha - G(c_0)}{G(c_0-) - G(c_0)}\,(G(c_0-) - G(c_0)) = \alpha.$$

Ainsi on a construit un test ϕ avec $c = c_0$ et $\gamma = \frac{\alpha - G(c_0)}{G(c_0-) - G(c_0)} \in [0,1]$ (avec la convention $0/0 = 0$) de niveau α pour $\{\theta_0\}$. Il reste à montrer qu'il est p.p. contre $\{\theta_1\}$. Soit ϕ^* un autre test de seuil α pour $\{\theta_0\}$, c'est-à-dire $\mathrm{E}_{\theta_0}[\phi^*(\mathbf{X})] \leqslant \alpha$.

On note

$$B^+ = \{\mathbf{x} \in \mathbb{R}^n : \phi(\mathbf{x}) - \phi^*(\mathbf{x}) > 0\} = \{\phi - \phi^* > 0\},$$

$$B^- = \{\mathbf{x} \in \mathbb{R}^n : \phi(\mathbf{x}) - \phi^*(\mathbf{x}) < 0\} = \{\phi - \phi^* < 0\}.$$

Alors, $B^+ \cap B^- = \emptyset$ et

$$B^+ \subset \{\phi = 1\} \cup \{\phi = \gamma\} = \{f_1 \geqslant c\, f_0\},$$

$$B^- \subset \{\phi = 0\} \cup \{\phi = \gamma\} = \{f_1 \leqslant c\, f_0\}.$$

Alors

$$(\#) \quad \int_{\mathbb{R}^n} (\phi(\mathbf{x}) - \phi^*(\mathbf{x}))\,(f_1(\mathbf{x}) - c\, f_0(\mathbf{x}))\,\mu(d\mathbf{x})$$

$$= \int_{B^+} (\phi(\mathbf{x}) - \phi^*(\mathbf{x}))\,(f_1(\mathbf{x}) - c\, f_0(\mathbf{x}))\,\mu(d\mathbf{x}) + \int_{B^-} (\phi(\mathbf{x}) - \phi^*(\mathbf{x}))\,(f_1(\mathbf{x}) - c\, f_0(\mathbf{x}))\,\mu(d\mathbf{x}) \geqslant 0.$$

Alors

$$\int_{\mathbb{R}^n} (\phi(\mathbf{x}) - \phi^*(\mathbf{x}))\, f_1(\mathbf{x})\mu(d\mathbf{x}) \geqslant c \int_{\mathbb{R}^n} (\phi(\mathbf{x}) - \phi^*(\mathbf{x}))\, f_0(\mathbf{x})\mu(d\mathbf{x}),$$

c'est-à-dire

$$\mathrm{E}_{\theta_1}[\phi(\mathbf{X})] - \mathrm{E}_{\theta_1}[\phi^*(\mathbf{X})] \geqslant c\,\mathrm{E}_{\theta_0}[\phi(\mathbf{X})] - \mathrm{E}_{\theta_1}[\phi^*(\mathbf{X})],$$

ou encore

$$\beta_\phi(\theta_1) - \beta_{\phi^*}(\theta_1) \geqslant c\,\alpha - \mathrm{E}_{\theta_0}[\phi^*(\mathbf{X})] \geqslant 0.$$

Ainsi $\beta_\phi(\theta_1) \geqslant \beta_{\phi^*}(\theta_1)$ pour tout autre test ϕ^* de seuil α pour $\{\theta_0\}$. □

Corollaire 2.1 *Soit ϕ le test défini par (1)-(2). Alors $\beta_\phi(\theta_1) \geqslant \alpha$. Ainsi ce test est sans biais contre $\{\theta_1\}$.*

Preuve. Le test constant $\phi^*(\mathbf{x}) \equiv \alpha$ est de niveau α pour $\{\theta_0\}$ et comme ϕ est p.p. contre $\{\theta_1\}$ on a $\beta_\phi(\theta_1) \geqslant \beta_{\phi^*}(\theta_1) = \alpha$. □

Remarques :

1) Le choix de c et γ est essentiellement unique. Si $c = c_0$ est un point de discontinuité de G alors c et γ sont uniquement déterminées. Si la droite passant par $(0, \alpha)$ est parallèle à l'axe des abscisses coupe le graphe de G en un seul point, alors $\gamma = 0$ et c est l'unique point pour lequel $G(c) = \alpha$. Si la même droite a une partie commune avec le graphe de G (sur $]b_1, b_2]$, disons), alors une fois de plus $\gamma = 0$ et on peut choisir n'importe quel c entre b_1 et b_2 sans modifier le niveau du test, car $\mathrm{P}_{\theta_0}(T(\mathbf{X}) \in]b_1, b_2]) == G(b_1) - G(b_2) = 0$.

Ainsi si G est continue on peut se limiter aux tests purs avec région critique $\{\mathbf{x} : \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} > c\}$.

2) La réciproque du théorème est vraie aussi : *S'il existe un test ψ satisfaisant (1)+(2) avec $c > 0$, alors tout test ϕ_0 p.p. contre $\{\theta_1\}$ de seuil α pour θ_0 est un test de niveau α pour θ_0 et il satisfait (1) exception sur un ensemble B tel que* $\mathrm{P}_{\theta_0}(\mathbf{X} \in B) = \mathrm{P}_{\theta_0}(\mathbf{X} \in B) = 0$.

Preuve : par la partie directe on sait que ψ est un test p.p. contre $\{\theta_1\}$ de seuil α pour θ_0, d'où $\beta_{\phi_0}(\theta_1) \geqslant \beta_\psi(\theta_1) \geqslant \beta_{\phi_0}(\theta_1)$, donc $\beta_{\phi_0}(\theta_1) = \beta_\psi(\theta_1)$. On a vu que, puisque ψ satisfait (1) et $0 \leqslant \phi_0 \leqslant 1$ que

$$0 \leqslant \beta_\psi(\theta_1) - \beta_{\phi_0}(\theta_1) - c[\beta_\psi(\theta_0) - \beta_{\phi_0}(\theta_0)]$$

et comme $c > 0$, $\beta_\psi(\theta_0) - \beta_{\phi_0}(\theta_0) \leqslant 0$. Mais ψ satisfait (2), donc $\beta_\psi(\theta_0) = \alpha$, d'où $\alpha \leqslant \beta_{\phi_0}(\theta_0) \leqslant \alpha$. Donc ϕ_0 est de niveau α pour $\{\theta_0\}$. De plus, comme dans la partie directe, on est en train d'intégrer un intégrant positif d'intégrale nulle (car l'inégalité (#) s'avère à être une égalité)donc ϕ_0 satisfait (1).

Exemples.
1) Supposons que la loi est $\mathcal{B}(1,\theta)$ et que $\theta_0 < \theta_1$. On voit que

$$\ln T(\mathbf{x}) = \left(\sum_{j=1}^n x_j\right) \ln \frac{\theta_1}{\theta_0} + \left(n - \sum_{j=1}^n x_j\right) \ln \frac{1-\theta_1}{1-\theta_0}.$$

La condition $T(\mathbf{x}) > c$ équivaut à $\sum_{j=1}^n x_j > c_0$, où

$$c_0 = \frac{\ln c - n \ln \frac{1-\theta_1}{1-\theta_0}}{\ln \frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}}.$$

Alors le test p.p. est

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } \sum_{j=1}^n x_j > c_0 \\ \gamma, & \text{si } \sum_{j=1}^n x_j = c_0 \\ 0, & \text{sinon,} \end{cases}$$

avec c_0 et γ données par

$$\mathrm{P}_{\theta_0}\left(\sum_{j=1}^n X_j > c_0\right) + \gamma\, \mathrm{P}_{\theta_0}\left(\sum_{j=1}^n X_j > c_0\right) = \alpha,$$

avec $\sum_{j=1}^n X_j \sim \mathcal{B}(n,\theta_i)$, $i = 0,1$.
2) Si la loi de l'échantillon est $\mathcal{P}(\theta)$ on trouve le même test avec

$$c_0 = \frac{\ln c e^{n(\theta_1-\theta_0)}}{\ln(\theta_1/\theta_0)},$$

et $\sum_{j=1}^n X_j \sim \mathcal{P}(n\theta_i)$, $i = 0,1$.
3) Si la loi est $\mathcal{N}(\theta,1)$, $?\theta_1 < \theta_0$

$$\ln T(\mathbf{x}) = \frac{1}{2}\sum_{j=1}^n \left[(x_j-\theta_0)^2 - (x_j-\theta_1)^2\right]$$

et $T(\mathbf{x}) > c$ si et seulement si $\bar{x} = \frac{1}{n}\sum_{j=1}^n x_j > c_0$, où

$$c_0 = \frac{1}{n}\left[\frac{\ln c}{\theta_1-\theta_0} + \frac{n(\theta_0+\theta_1)}{2}\right].$$

Le test est pur

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } \bar{x} > c_0 \\ 0, & \text{sinon,} \end{cases}$$

où c_0 donnée par $\mathrm{P}_{\theta_0}\left(\bar{X} > c_0\right) = \alpha$ et $\bar{X} \sim \mathcal{N}(\theta_i, \frac{1}{n})$, $i = 0, 1$.
4) Enfin si la loi est $\mathcal{N}(0, \theta)$, ?$\theta_1 < \theta_0$

$$\ln T(\mathbf{x}) = \frac{\theta_1 - \theta_0}{2\theta_0\theta_1}\left(\sum_{j=1}^{n} x_j^2\right) + \frac{1}{2}\ln\frac{\theta_0}{\theta_1}.$$

La condition $T(\mathbf{x}) > c$ si et seulement si $\sum_{j=1}^{n} x_j^2 > c_0$, où

$$c_0 = \frac{2\theta_0\theta_1}{\theta_1 - \theta_0}\ln\left(c\sqrt{\frac{\theta_1}{\theta_0}}\right).$$

Le test est pur

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } \sum_{j=1}^{n} x_j^2 > c_0 \\ 0, & \text{sinon,} \end{cases}$$

où c_0 donnée par $\mathrm{P}_{\theta_0}\left(\sum_{j=1}^{n} X_j^2 > c_0\right) = \alpha$. Ici $\frac{1}{\theta_i}\sum_{j=1}^{n} X_j^2 \sim \chi^2(n)$, $i = 0, 1$.

2.2.2 Tests u.p.p. pour certains hypothèses composites

Dans la plupart des problèmes d'intérêt pratique au moins une des hypothèses est composite. On va montrer que pour certaines familles de lois et certaines hypothèses il existe des tests u.p.p.

Définition 2.7 *Soit* $\mathbf{X}$ *un n-échantillon de densité* $f_\theta(\cdot)$, $\theta \in \Theta \subset \mathbb{R}$. *Cette famille de lois est* **à rapport de vraisemblance monotone en** S *si l'ensemble* $D := \{\mathbf{x} \in \mathbb{R}^n : f_\theta(\mathbf{x}) > 0\}$ *ne dépend pas de* θ *et il existe une fonction mesurable* $S : \mathbb{R}^n \to \mathbb{R}$ *telle que lorsque* $\theta, \theta' \in \Theta$ *avec* $\theta < \theta'$, *on a*

$$f_\theta(\mathbf{x}) \neq f_{\theta'}(\mathbf{x}),\ \forall \mathbf{x} \in D$$

et

$$R(\mathbf{x}, \theta, \theta') = \frac{f_{\theta'}(\mathbf{x})}{f_\theta(\mathbf{x})} = \mathtt{F}(S(\mathbf{x})) \text{ est une fonction monotone de } S(\mathbf{x}).$$

Notons que le rapport de vraisemblance est bien défini sauf sur un ensemble $N \subset \mathbb{R}^n$ *tel que* $\mathrm{P}_\theta(\mathbf{X} \in N) = 0$, *pour tout* $\theta \in \Theta$. *On travaillera toujours en dehors d'un tel ensemble.*

Proposition 2.1 *Soit la famille de lois de densité de type exponentiel*

$$f_\theta(x_1) = h(x_1)\exp\{a(\theta)U(x_1) + V(\theta)\}$$

et supposons que $a(\cdot)$ *est croissante. Alors la famille est à rapport de vraisemblance monotone croissant en* S, *où* $S(\mathbf{x}) = \sum_{j=1}^{n} U(x_j)$. *Si* $a(\cdot)$ *est décroissante alors le rapport de vraisemblance est décroissant en* S.

Preuve. On a

$$f_\theta(\mathbf{x}) = h(\mathbf{x}) \exp\{a(\theta)S(\mathbf{x}) + nV(\theta)\}, \; h(\mathbf{x}) = h(x_1)\dots h(x_n).$$

Sur l'ensemble de points $\mathbf{x}$ où $h(\mathbf{x}) > 0$ (de P_θ-probabilité 1 pour tout θ) on a

$$R(\mathbf{x}, \theta, \theta') = \frac{f_{\theta'}(\mathbf{x})}{f_\theta(\mathbf{x})} = e^{(a(\theta')-a(\theta))S(\mathbf{x})} e^{n(V(\theta')-V(\theta))}.$$

La conclusion s'ensuit facilement. □

Théorème 2.2 *Soit* $\mathbf{X}$ *un* n*-échantillon de densité* $f_\theta(\cdot)$, $\theta \in \Theta \subset \mathbb{R}$. *Supposons qu'il s'agit d'une famille à rapport de vraisemblance croissant en* S. *Soit* $\theta_0 \in \Theta$ *fixé et* $\Theta_0 = \{\theta \in \Theta : \theta \leqslant \theta_0\}$. *Pour tester*

$$(\mathrm{H}) : \theta \leqslant \theta_0 \;\; \textit{contre} \;\; (\mathrm{A}) : \theta > \theta_0$$

au seuil $\alpha \in]0,1[$ *pour* $\{\theta \leqslant \theta_0\}$ *il y a un test u.p.p. contre* $\{\theta > \theta_0\}$ *dans la classe des tests de seuil* α *pour* $\{\theta \leqslant \theta_0\}$ *donné par*

$$\phi(\mathbf{x}) = \begin{cases} 1, & \textit{si } S(\mathbf{x}) > c \\ \gamma, & \textit{si } S(\mathbf{x}) = c \\ 0, & \textit{sinon,} \end{cases} \tag{3}$$

où les constantes $\gamma \in [0,1]$ *et* $c > 0$ *sont déterminées par*

$$\mathrm{E}_{\theta_0}[\phi(\mathbf{X})] = \mathrm{P}_{\theta_0}(S(\mathbf{X}) > c) + \gamma\, \mathrm{P}_{\theta_0}(S(\mathbf{X}) = c) = \alpha. \tag{4}$$

Si le rapport de vraisemblance est décroissant en S *on renverse les inégalités en (3)-(4).*

Preuve. *1er pas.* Soient $\theta' < \theta''$ dans Θ et on considère le problème de tester au niveau α

$$(\mathrm{H}') : \theta = \theta' \;\; \text{contre} \;\; (\mathrm{H}'') : \theta = \theta''.$$

Le lemme fondamental de Neyman-Pearson assure que le test

$$\phi'''(\mathbf{x}) = \begin{cases} 1, & \text{si } f_{\theta''}(\mathbf{x}) > k\, f_{\theta'}(\mathbf{x}) \\ \gamma''', & \text{si } f_{\theta''}(\mathbf{x}) = k\, f_{\theta'}(\mathbf{x}) \\ 0, & \text{sinon,} \end{cases}$$

avec $\mathrm{E}_{\theta'}[\phi'''(\mathbf{X})] = \alpha$ est (u.) p.p. contre $\{\theta''\}$ au seuil α pour $\{\theta'\}$. Mais $\frac{f_{\theta''}(\mathbf{x})}{f_{\theta'}(\mathbf{x})} = \mathtt{F}(S(\mathbf{x}))$ avec $\mathtt{F} : \mathbb{R} \to \mathbb{R}$ croissante. Alors

$$\mathtt{F}(S(\mathbf{x})) > k \Leftrightarrow S(\mathbf{x}) > c'''(:= \mathtt{F}^{-1}(k))$$

donc

$$\phi'''(\mathbf{x}) = \begin{cases} 1, & \text{si } S(\mathbf{x}) > c''' \\ \gamma''', & \text{si } S(\mathbf{x}) = c''' \\ 0, & \text{sinon,} \end{cases}$$

et $\mathrm{E}_{\theta'}[\phi'''(\mathbf{X})] = \mathrm{P}_{\theta'}(S(\mathbf{X}) > c''') + \gamma'''\mathrm{P}_{\theta'}(S(\mathbf{X}) = c''') = \alpha$. De plus $\beta_{\phi'''}(\theta'') \geqslant \alpha$.
2ème pas. On prend $\theta' = \theta_0$ et $\theta'' > \theta' = \theta_0$ arbitraire. On teste

$$(\mathrm{H}_0) : \theta = \theta_0 \ \text{ contre } \ (\mathrm{H}'') : \theta = \theta''.$$

On voit que ϕ de l'énoncé coïncide avec ϕ''' avec $c = c'''$ et $\gamma = \gamma'''$. Donc $\alpha = \mathrm{E}_{\theta'}[\phi'''(\mathbf{X})] = \mathrm{E}_{\theta_0}[\phi(\mathbf{X})]$ c'est-à-dire ϕ est de niveau α pour $\{\theta_0\}$. De plus ϕ est (u.) p.p. contre $\{\theta''\}$ au seuil α pour $\{\theta_0\}$. Comme θ'' est arbitraire, on déduit que ϕ est u.p.p. contre $\{\theta > \theta_0\}$ au seuil α pour $\{\theta_0\}$.
3ème pas. On prend $\theta'' = \theta_0$ et $\theta' < \theta'' = \theta_0$ arbitraire. On teste

$$(\mathrm{H}') : \theta = \theta' \ \text{ contre } \ (\mathrm{H}_0) : \theta = \theta_0.$$

On voit que ϕ de l'énoncé coïncide avec ϕ''' avec $c = c'''$ et $\gamma = \gamma'''$ et il est de niveau $\alpha(\theta') = \mathrm{E}_{\theta'}[\phi(\mathbf{X})]$ pour $\{\theta'\}$. Comme $\beta_{\phi'''}(\theta'') \geqslant \alpha$, on a que $\mathrm{E}_{\theta_0}[\phi(\mathbf{X})] \geqslant \alpha(\theta')$, autrement dit

$$\alpha = \mathrm{E}_{\theta_0}[\phi(\mathbf{X})] \geqslant \alpha(\theta') = \mathrm{E}_{\theta'}[\phi(\mathbf{X})]$$

(on voit aussi la monotonie de la fonction puissance). Donc ϕ est de seuil α pour $\{\theta'\}$. De plus ϕ est (u.) p.p. contre $\{\theta_0\}$ au seuil α pour $\{\theta'\}$. Comme θ' est arbitraire, on déduit que ϕ est u.p.p. contre $\{\theta_0\}$ au seuil α pour $\{\theta < \theta_0\}$.
4ème pas. Soit

$$\mathcal{C} = \{\text{tous les tests de seuil } \alpha \text{ pour } \theta \leqslant \theta_0\},$$

$$\mathcal{C}_0 = \{\text{tous les tests de seuil } \alpha \text{ pour tester } \theta = \theta_0\}.$$

Alors clairement, $\mathcal{C} \subset \mathcal{C}_0$, car tout test ψ avec $\sup_{\theta \leqslant \theta_0} \mathrm{E}_\theta[\psi(\mathbf{X})] \leqslant \alpha$ satisfait $\mathrm{E}_{\theta_0}[\psi(\mathbf{X})] \leqslant \alpha$. Par le *3ème pas* ϕ de l'énoncé est dans $\mathcal{C}$. Par le *2ème pas* ϕ u.p.p. contre $\{\theta > \theta_0\}$ parmi les tests de $\mathcal{C}_0$, donc il sera u.p.p. contre $\{\theta > \theta_0\}$ parmi les tests d'une sous-classe de $\mathcal{C}_0$, donc de $\mathcal{C}$. Le résultat s'ensuit. □

Remarque : Pour le problème symétrique où $\Theta_0 = \{\theta \in \Theta : \theta \geqslant \theta_0\}$, sous les hypothèses du Théorème 2.2, il existe un test u.p.p. pour tester (H) : $\theta \in \Theta_0$ contre (A) : $\theta \in \Theta_0^c$. Le test est encore donné par (3)-(4) si le rapport de vraisemblance est décroissante en $S(\mathbf{x})$ et avec des inégalités renversées si le rapport de vraisemblance est croissante en $S(\mathbf{x})$.

Corollaire 2.2 *Soit* $\mathbf{X}$ *un* n*-échantillon de densité de type exponentiel*

$$f_\theta(x_1) = h(x_1) \exp\{a(\theta)U(x_1) + V(\theta)\}$$

et supposons que $a(\cdot)$ *est strictement monotone. Alors pour tester* (H) : $\theta \in \Theta_0 = \{\theta \leqslant \theta_0\}$ *contre* (A) : $\theta \in \Theta_0^c$ *au niveau* α *il existe un test u.p.p. contre* Θ_0^c *dans la classe des tests* *de seuil α pour Θ_0. Ce test est donné par (3)-(4) si $a(\cdot)$ est croissante et par (3)-(4) avec des inégalités renversées si $a(\cdot)$ est décroissante.*

Pour tester (H) : $\theta \in \Theta_0 = \{\theta \geqslant \theta_0\}$ *contre* (A) : $\theta \in \Theta_0^c$ *au niveau α il existe un test u.p.p. Θ_0^c dans la classe des tests de seuil α pour Θ_0. Ce test est donné par (3)-(4) si $a(\cdot)$ est décroissante et par (3)-(4) avec des inégalités renversées si $a(\cdot)$ est croissante.*

Remarque : On peut aussi montrer que la fonction puissance pour le problème discuté au Théorème 2.2, $\beta(\theta) = \mathrm{E}_\theta\left[\phi(\mathbf{X})\right]$ étendue pour tout $\theta \in \Theta$ est une fonction croissante pour ces θ pour lesquels cette fonction est plus petite que 1.

Théorème 2.3 *Soit $\mathbf{X}$ un n-échantillon de densité de type exponentiel*

$$f_\theta(x_1) = h(x_1)\exp\left\{a(\theta)U(x_1) + V(\theta)\right\},\ \theta \in \Theta \subset \mathbb{R}.$$

Supposons qu'il s'agit d'une famille à rapport de vraisemblance croissante en S. (autrement dit, supposons que $a(\cdot)$ est strictement croissante). Soit $\theta_1, \theta_2 \in \Theta$ fixés et $\theta_1 < \theta_2$. Pour tester

$$\text{(H)} : \theta \in \Theta_0 = \{\theta \in \Theta : \theta \leqslant \theta_1 \ \ \textit{ou}\ \ \theta \geqslant \theta_2\} \ \ \textit{contre}\ \ \text{(A)} : \theta \in \Theta_0^c = \{\theta_1 < \theta < \theta_2\}$$

au niveau $\alpha \in]0,1[$, le test donné par

$$\phi(\mathbf{x}) = \begin{cases} 1, & \textit{si}\ \ c_1 < S(\mathbf{x}) < c_2 \\ \gamma_i, & \textit{si}\ S(\mathbf{x}) = c_i,\ i = 1,2 \\ 0, & \textit{sinon,} \end{cases} \tag{5}$$

où les constantes γ_1, γ_2 et c_1, c_2 sont déterminées par

$$\mathrm{E}_{\theta_i}\left[\phi(\mathbf{X})\right] = \mathrm{P}_{\theta_i}(c_1 < S(\mathbf{X}) < c_2) + \gamma_1 \mathrm{P}_{\theta_i}(S(\mathbf{X}) = c_1) + \gamma_2 \mathrm{P}_{\theta_i}(S(\mathbf{X}) = c_2) = \alpha,\ i = 1,2, \tag{6}$$

est u.p.p. contre Θ_0^c au seuil α pour Θ_0. Si le rapport de vraisemblance est décroissant en S (autrement dit, si $a(\cdot)$ est décroissante), alors dans (5) on prend

$$\phi(\mathbf{x}) = 1 \ \textit{si}\ S(\mathbf{x}) < c_1 \ \ \textit{ou}\ \ c_2 < S(\mathbf{x}).$$

Remarque : On peut montrer que pour ce problème la fonction puissance $\beta(\theta) = \mathrm{E}_\theta\left[\phi(\mathbf{X})\right]$ étendue à tout Θ, est croissante pour $\theta \leqslant \theta_0$ et décroissante pour $\theta \geqslant \theta_0$, pour un certain $\theta_1 < \theta_0 < \theta_2$.

Pour prouver le Théorème 2.3 on a besoin d'un résultat plus général que le lemme fondamental de Neyman-Pearson.

Soient $\mathrm{P}_1, \ldots, \mathrm{P}_{m+1}$ des lois et on suppose que leurs densités sont $f_1, \ldots, f_{m+1}$ par rapport à $\mu = \sum_{k=1}^{m+1} \mathrm{P}_k$. Soit le convexe compact de $\mathbb{R}^{m+1}$

$$C_{m+1} = \{\mathrm{E}_\ell[\phi(\mathbf{X})] : \ell = 1, \ldots, m+1, \phi \ \text{fonction à valeurs dans}\ [0,1]\}$$

et de la même façon C_m.

Théorème 2.4 *(lemme de Neyman-Pearson généralisé)*
1) Il existe un test ϕ *maximisant* $\mathrm{E}_{m+1}[\phi(\mathbf{X})]$ *sous les contraintes*

$$\mathrm{E}_\ell[\phi(\mathbf{X})] = c_\ell,\ \ell = 1, \ldots, m,\ \mathbf{c} = (c_1, \ldots, c_m) \in C_m. \tag{7}$$

2) Si ϕ *vérifie les contraintes (7) et est de type (N-P), c'est-à-dire*

$$\phi(\mathbf{x}) = \begin{cases} 1 & \textit{sur } \{f_{m+1} > \sum_{\ell=1}^m k_\ell f_\ell\} \\ 0 & \textit{sur } \{f_{m+1} < \sum_{\ell=1}^m k_\ell f_\ell\}, \end{cases} \tag{8}$$

il maximise $\mathrm{E}_{m+1}[\phi(\mathbf{X})]$ *sous les contraintes (7).*
3) Si ϕ *vérifie les contraintes (7) et est de type (N-P) avec* $k_\ell \geqslant 0$*, il maximise* $\mathrm{E}_{m+1}[\phi(\mathbf{X})]$ *sous les contraintes* $\mathrm{E}_\ell[\phi(\mathbf{X})] \leqslant c_\ell,\ \ell = 1, \ldots, m.$
4) Si $\mathbf{c}$ *est un point intérieur de* C_m*, il existe un test de type (N-P) vérifiant les contraintes (7), et tout test maximisant* $\mathrm{E}_{m+1}[\phi(\mathbf{X})]$ *avec les contraintes (7) est de type (N-P).*

Ce résultat sera admis. Donnons une idée de preuve d'une partie des affirmations. Soient $c_1, \ldots, c_m$ constantes et soit $\mathcal{C}$ la classe des fonctions $0 \leqslant \phi \leqslant 1$ telles que $\int \phi(\mathbf{x}) f_\ell(\mathbf{x}) \mu^n(d\mathbf{x}) = c_\ell$. Soit ϕ^* telle que

$$\phi^*(\mathbf{x}) = \begin{cases} 1 & \text{si } f_{m+1}(\mathbf{x}) > \sum_{\ell=1}^m k_\ell f_\ell(\mathbf{x}) \\ 0 & \text{si } f_{m+1}(\mathbf{x}) < \sum_{\ell=1}^m k_\ell f_\ell(\mathbf{x}) \end{cases}$$

Montrons que ϕ^* maximise $\int \phi(\mathbf{x}) f_{m+1}(\mathbf{x}) \mu^n(d\mathbf{x})$ sur $\mathcal{C}$. Soit ϕ dans $\mathcal{C}$. On a

$$\begin{aligned} 0 &\leqslant \int [\phi^*(\mathbf{x}) - \phi(\mathbf{x})][f_{m+1}(\mathbf{x}) - \sum_{\ell=1}^m k_\ell f_\ell(\mathbf{x})] \mu^n(d\mathbf{x}) \\ &= \int \phi^*(\mathbf{x}) f_{m+1}(\mathbf{x}) \mu^n(d\mathbf{x}) - \int \phi(\mathbf{x}) f_{m+1}(\mathbf{x}) \mu^n(d\mathbf{x}) \\ &- \sum_{\ell=1}^m k_\ell [\int \phi^*(\mathbf{x}) f_\ell(\mathbf{x}) \mu^n(d\mathbf{x}) - \int \phi(\mathbf{x}) f_\ell(\mathbf{x}) \mu^n(d\mathbf{x})] \\ &= \int \phi^*(\mathbf{x}) f_{m+1}(\mathbf{x}) \mu^n(d\mathbf{x}) - \int \phi(\mathbf{x}) f_{m+1}(\mathbf{x}) \mu^n(d\mathbf{x}), \end{aligned}$$

d'où la conclusion.
Remarque : On introduit la *diagramme des risques* $\mathcal{R}$ comme étant l'ensemble de $\mathbb{R}^2$ de coordonnées $(\alpha_\phi, \beta_\phi)$ quand ϕ parcourt l'ensemble des tests. On a $\mathcal{R} \subset [0,1] \times [0,1]$, $(0,1), (1,0) \in \mathcal{R}$ seuls points d'intersection avec le carré $[0,1] \times [0,1]$, $\mathcal{R}$ est symétrique par rapport à $(\frac{1}{2}, \frac{1}{2})$ et est convexe, fermé compact.

Preuve du Théorème 2.3. *(facultatif)* Le point (α, α) est intérieur au diagramme des risques correspondant à P_{θ_1} et P_{θ_2}. Par le dernier point du lemme généralisé de

Neyman-Pearson, tout test ϕ_α maximisant $\mathrm{E}_{\theta'}[\phi(\mathbf{X})]$, $\theta_1 < \theta' < \theta_2$ sous les contraintes $\mathrm{E}_{\theta_1}[\phi(\mathbf{X})] = \mathrm{E}_{\theta_2}[\phi(\mathbf{X})] = \alpha$ est de type

$$\phi_\alpha(\mathbf{x}) = 1 \text{ si } f_{\theta'} > k_1 f_{\theta_1} + k_2 f_{\theta_2},$$

c'est-à-dire si

$$e^{nV(\theta')} e^{a(\theta')S(\mathbf{x})} > k_1 e^{nV(\theta_1)} e^{a(\theta_1)S(\mathbf{x})} + k_2 e^{nV(\theta_2)} e^{a(\theta_2)S(\mathbf{x})}$$

ou encore

$$1 > d_1 e^{b_1 S(\mathbf{x})} + d_2 e^{b_2 S(\mathbf{x})},$$

avec $b_1 = a(\theta_1) - a(\theta') < 0$ et $b_2 = a(\theta_2) - a(\theta') > 0$. On peut vérifier aisément que $d_1, d_2 > 0$. D'où l'inégalité précédente est équivalente à $c_1 < S(\mathbf{x}) < c_2$.

Si on sait que le test ϕ_α est de niveau α, d'après *3)* du lemme généralisé ϕ_α maximise $\mathrm{E}_\theta[\phi(\mathbf{X})]$ sous les contraintes $\mathrm{E}_{\theta_i}[\phi(\mathbf{X})] \leqslant \alpha$, $i = 1, 2$, donc sous contrainte $\mathrm{E}_\theta[\phi(\mathbf{X})] \leqslant \alpha$, $\forall \theta \in \Theta_0$.

Reste à voir que le test ϕ_α est de niveau α. D'après *2)* du lemme généralisé on sait qu'un test ϕ'_α vérifiant les contraintes $\mathrm{E}_{\theta_i}[\phi'_\alpha(\mathbf{X})] = \alpha$, $i = 1, 2$, avec $\phi'_\alpha(\mathbf{x}) = 1$ si $f_{\theta''} < k'_1 f_{\theta_1} + k'_2 f_{\theta_2}$, minimise $\mathrm{E}_{\theta''}[\phi(\mathbf{X})]$ sous les mêmes contraintes (car on applique cette partie *2)* à $1 - \phi'_\alpha$). On voit alors comme ci-dessus que $k'_1 > 0$ et $k'_2 < 0$ et donc que le test ϕ_α vérifie les condition imposées à ϕ'_α. Donc $\mathrm{E}_{\theta''}[\phi_\alpha(\mathbf{X})]$ minimise $\mathrm{E}_{\theta''}[\phi(\mathbf{X})]$ sous les contraintes. On considère le test trivial $\phi^* \equiv \alpha$, donc $\mathrm{E}_{\theta''}[\phi_\alpha(\mathbf{X})] \leqslant \alpha$, $\forall \theta'' \in \Theta_0$. $\square$

Exemples : Dans tous les exemples suivants on va tester

$$(\mathrm{H}) : \theta \in \Theta_0 = \{\theta \in \Theta : \theta \leqslant \theta_0\} \text{ contre } (\mathrm{A}) : \theta \in \Theta_0^c$$

ou

$$(\mathrm{H}') : \theta \in \Theta_0 = \{\theta \in \Theta : \theta \leqslant \theta_1 \text{ ou } \theta \geqslant \theta_2\} \text{ contre } (\mathrm{A}') : \theta \in \Theta_0^c$$

où $\theta_0, \theta_1, \theta_2 \in \Theta$, $\theta_1 < \theta_2$. Le niveau sera $\alpha \in]0, 1[$.

1. Soit $\mathbf{X}$ un n-échantillon de loi $\mathcal{B}(1, \theta)$, $\theta \in \Theta =]0, 1[$. Dans ce cas

$$S(\mathbf{x}) = \sum_{j=1}^{n} x_j \text{ et } a(\theta) = \ln \frac{\theta}{1-\theta}.$$

De plus $a(\cdot)$ est croissante. Donc par le Corollaire 2.2, le test pour tester (H) est donné par :

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } \sum_{j=1}^{n} x_j > c \\ \gamma, & \text{si } \sum_{j=1}^{n} x_j = c \\ 0, & \text{sinon,} \end{cases}$$

où c, γ sont données par

$$\mathrm{P}_{\theta_0}(S > c) + \gamma \mathrm{P}_{\theta_0}(S = c) = \alpha, \text{ avec } S \sim \mathcal{B}(n, \theta).$$

Par le Théorème 2.3, pour tester (H′) le test est

$$\phi'(\mathbf{x}) = \begin{cases} 1, & \text{si } c_1 < \sum_{j=1}^n x_j < c_2 \\ \gamma_i, & \text{si } \sum_{j=1}^n x_j = c_i, i = 1, 2 \\ 0, & \text{sinon,} \end{cases}$$

où

$$\mathrm{P}_{\theta_i}(c_1 < S < c_2) + \gamma_1 \mathrm{P}_{\theta_i}(S = c_1) + \gamma_2 \mathrm{P}_{\theta_i}(S = c_2) = \alpha,\ i = 1, 2.$$

2. Soit $\mathbf{X}$ un n-échantillon de loi $\mathcal{P}(\theta)$, $\theta \in \Theta =]0, \infty[$. Dans ce cas

$$S(\mathbf{x}) = \sum_{j=1}^n x_j \text{ et } a(\theta) = \ln \theta \text{ est croissante.}$$

Les tests sont identiques avec $S \sim \mathcal{P}(n\theta)$.
3. Soit $\mathbf{X}$ un n-échantillon de loi $\mathcal{N}(\theta, \sigma^2)$ avec σ^2 connu. Dans ce cas

$$S(\mathbf{x}) = \sum_{j=1}^n x_j \text{ et } a(\theta) = \frac{1}{\sigma^2}\theta \text{ est croissante.}$$

Le test pour (H) est

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{si } \bar{x} \geqslant c \\ 0, & \text{sinon,} \end{cases}$$

où $\bar{x} = \frac{1}{n}\sum_{j=1}^n x_j$ et c est donnée par

$$\mathrm{P}_{\theta_0}(\bar{X} \geqslant c) = \alpha,\ \bar{X} \sim \mathcal{N}(\theta, \frac{\sigma^2}{n}).$$

La puissance du test est

$$\beta_\phi(\theta) = 1 - \Phi\left(\frac{\sqrt{n}(c - \theta)}{\sigma}\right).$$

Par exemple pour $\sigma = 2$, $\theta_0 = 20$, $\alpha = 0,05$ et $n = 25$ on trouve $c = 20,66$ et la puissance du test pour $\theta_1 = 21$ est $\beta_\phi(21) = 0,8023$.
Pour tester (H′) on prend

$$\phi'(\mathbf{x}) = \begin{cases} 1 & \text{si } c_1 \leqslant \bar{x} \leqslant c_2 \\ 0, & \text{sinon,} \end{cases}$$

où c_1, c_2 sont données par

$$\mathrm{P}_{\theta_i}(c_1 \leqslant \bar{X} \leqslant c_2) = \alpha,\ i = 1, 2.$$

La puissance est

$$\beta_{\phi'}(\theta) = \Phi\left(\frac{\sqrt{n}(c_2 - \theta)}{\sigma}\right) - \Phi\left(\frac{\sqrt{n}(c_1 - \theta)}{\sigma}\right)$$

Par exemple pour $\sigma = 2$, $\theta_1 = -1$, $\theta_2 = 1$, $\alpha = 0,05$ et $n = 25$ on trouve $c_1 = -0,344$, $c_2 = 0,344$ et la puissance du test pour $\theta = 0$ est $\beta_{\phi'}(0) = 0,61$.
4. Soit $\mathbf{X}$ un n-échantillon de loi $\mathcal{N}(m, \theta)$ avec m connu. Dans ce cas

$$S(\mathbf{x}) = \sum_{j=1}^{n} (x_j - m)^2 \text{ et } a(\theta) = -\frac{1}{2\theta} \text{ est croissante.}$$

Le test pour (H) est

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{si } \sum_{j=1}^{n} (x_j - m)^2 \geqslant c \\ 0, & \text{sinon,} \end{cases}$$

où

$$\mathrm{P}_{\theta_0}(\sum_{j=1}^{n} (X_j - m)^2 \geqslant c) = \alpha, \ \frac{1}{\theta} \sum_{j=1}^{n} (X_j - m)^2 \sim \chi^2(n).$$

La puissance est $\beta_\phi(\theta) = 1 - \mathrm{P}(Y \leqslant c/\theta)$, $Y \sim \chi^2(n)$, donc ne dépend pas de m. Par exemple pour $\theta_0 = 4$, $\alpha = 0,05$ et $n = 25$ on trouve $c = 150,609$ et pour $\theta = 12$ la puissance est $\beta_\phi(12) = 0,98$.
Le test pour (H$'$) est

$$\phi'(\mathbf{x}) = \begin{cases} 1 & \text{si } c_1 \leqslant \sum_{j=1}^{n} (x_j - m)^2 \leqslant c_2 \\ 0, & \text{sinon,} \end{cases}$$

où

$$\mathrm{P}_{\theta_i}(c_1 \leqslant \sum_{j=1}^{n} (X_j - m)^2 \leqslant c_2) = \alpha, \ i = 1, 2.$$

La puissance est

$$\beta_{\phi'}(\theta) = \mathrm{P}\left(Y \leqslant \frac{c_2}{\theta}\right) - \mathrm{P}\left(Y \leqslant \frac{c_1}{\theta}\right)$$

Par exemple pour $\theta_1 = 1$, $\theta_2 = 3$, $\alpha = 0,01$ et $n = 25$ on a

$$\mathrm{P}\left(Y \leqslant c_2\right) - \mathrm{P}\left(Y \leqslant c_1\right) = 0,01, \ \mathrm{P}\left(Y \leqslant \frac{c_2}{3}\right) - \mathrm{P}\left(Y \leqslant \frac{c_1}{3}\right) = 0,01$$

et c_1, c_2 sont déterminées (par plusieurs essais).

2.2.3 Tests u.p.p.s.b. pour certains hypothèses composites

Il n'est pas difficile de voir qu'il n'existe pas de test u.p.p. pour tester

$$(\mathrm{H}) : \theta \in \Theta_0 = \{\theta \in \Theta : \theta_1 < \theta < \theta_2\} \text{ contre } (\mathrm{A}) : \theta \in \Theta_0^c$$

ou

$$(\mathrm{H}_0) : \theta = \theta_0 \text{ contre } (\mathrm{A}'') : \theta \neq \theta_0.$$

Par exemple pour ce dernier cas, on voit que le test donné par (3)-(4) du Théorème 2.2 est u.p.p. pour $\theta > \theta_0$ mais il moins bon que le test trivial $\phi^*(\mathbf{x}) = \alpha$, pour $\theta < \theta_0$. Ainsi il n'existe pas de test qu'il soit u.p.p. pour tout $\theta \neq \theta_0$. Il faut réduire la classe des tests : sans biais.

Théorème 2.5 *Soit* $\mathbf{X}$ *un n-échantillon de densité de type exponentiel*

$$f_\theta(x_1) = h(x_1)\exp\left\{\theta U(x_1) + V(\theta)\right\}, \, \theta \in \Theta \subset \mathbb{R}.$$

On fixe $\theta_0, \theta_1, \theta_2 \in \Theta$, $\theta_1 < \theta_2$ *et on pose* $\Theta_{12} = \{\theta \in \Theta : \theta_1 \leqslant \theta \leqslant \theta_2\}$ *et* $\Theta_0 = \{\theta_0\}$. *Alors pour tester l'hypothèse* (H) : $\theta \in \Theta_{12}$ *contre* (A) : $\theta \in \Theta_{12}^c$ *ou pour tester* (H_0) : $\theta \in \Theta_0$ *contre* (A_0) : $\theta \in \Theta_0^c$ *au niveau* α, *il existe des tests u.p.p.s.b. donnés par*

$$\phi(\mathbf{x}) = \begin{cases} 1, & \textit{si } S(\mathbf{x}) < c_1 \textit{ ou } S(\mathbf{x}) > c_2 \\ \gamma_i, & \textit{si } S(\mathbf{x}) = c_i,\, i = 1, 2\, (c_1 < c_2) \\ 0, & \textit{sinon,} \end{cases} \tag{9}$$

où les constantes γ_1, γ_2 *et* c_1, c_2 *sont déterminées par*

$$\mathrm{E}_{\theta_i}\left[\phi(\mathbf{X})\right] = \alpha,\, i = 1, 2 \textit{ pour } (\mathrm{H}) \tag{10}$$

et

$$\mathrm{E}_{\theta_0}\left[\phi(\mathbf{X})\right] = \alpha,\, \mathrm{E}_{\theta_0}\left[S(\mathbf{X})\phi(\mathbf{X})\right] = \alpha \mathrm{E}_{\theta_0}[S(\mathbf{X})] \textit{ pour } (\mathrm{H}_0). \tag{11}$$

On rappelle que $S(\mathbf{x}) = \sum_{j=1}^n U(x_j)$.

Remarque : On peut montrer que la fonction puissance $\beta_\phi = \mathrm{E}_\theta[\phi(Z)]$, $\theta \in \Theta$ est décroissante pour $\theta \leqslant \theta_0$ et croissante pour $\theta \geqslant \theta_0$, pour un certain $\theta_0 \in]\theta_1, \theta_2[$ (pour (H)) ou pour θ_0 (pour (H_0)).
Remarque : On peut se demander si les lois classiques comme binomiale, de Poisson ou gaussienne sont de la forme de l'énoncé. Il suffit de faire une reparamétrisation pour les mettre sous cette forme. Sont des familles exponentielles :

$$f_\theta(x) = h(x) \cdot \exp\left\{a(\theta)U(x) + V(\theta)\right\}$$

i) pour $\mathcal{B}(1, \theta)$, $a(\theta) = \ln(\theta/(1-\theta)]$; on pose $\tau = \ln[\theta/(1-\theta)]$. Ainsi la densité est de la forme du théorème. On a $\theta = e^\tau/(1+e^\tau)$ et les hypothèses $\theta_1 \leqslant \theta \leqslant \theta_2$, $\theta = \theta_0$ deviennent $\tau_1 \leqslant \tau \leqslant \tau_2$, $\tau = \tau_0$, où $\tau_i = \ln[\theta_i/(1-\theta_i)]$, $i = 0, 1, 2$.
ii) pour $\mathcal{P}(\theta)$, $a(\theta) = \ln\theta$; on pose $\tau = \ln\theta$ et on retrouve la forme désirée. On a $\theta = e^\tau$ et les hypothèses deviennent $\tau_1 \leqslant \tau \leqslant \tau_2$, $\tau = \tau_0$, où $\tau_i = \ln\theta_i$, $i = 0, 1, 2$.
iii) pour $\mathcal{N}(\theta, \sigma^2)$ avec σ connu, $a(\theta) = (1/\sigma^2)\theta$. Le facteur $1/\sigma^2$ peut être absorbé dans $U(x)$.
iv) pour $\mathcal{N}(m, \theta)$ avec m connu, $a(\theta) = -1/(2\theta)$; on pose $\tau = -1/(2\theta)$ et on retrouve la forme désirée. On a $\theta = -1/(2\tau)$ à nouveau les hypothèses deviennent $\tau_1 \leqslant \tau \leqslant \tau_2$, $\tau = \tau_0$, où $\tau_i = -1/(2\theta_i)$, $i = 0, 1, 2$.

On admet le résultat du théorème mais on va l'illustrer avec l'exemple suivant :

Exemple : Soit $\mathbf{X}$ un n-échantillon de loi $\mathcal{N}(\theta, \sigma^2)$ avec σ connu. On veut tester (H_0) : $\theta = \theta_0$ contre (A_0) : $\theta \neq \theta_0$. Ici

$$U(x) = \frac{1}{\sigma^2}x,$$

donc

$$S(\mathbf{x}) = \sum_{j=1}^{n} U(x_j) = \frac{1}{\sigma^2}\sum_{j=1}^{n} x_j = \frac{n}{\sigma^2}\bar{x}.$$

Le test u.p.p.s.b. est donné par

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } \frac{n}{\sigma^2}\bar{x} < c_1 \text{ ou } \frac{n}{\sigma^2}\bar{x} > c_2 \\ 0, & \text{sinon,} \end{cases}$$

où c_1, c_2 sont données par

$$\mathrm{E}_{\theta_0}[\phi(\mathbf{X})] = \alpha,\ \mathrm{E}_{\theta_0}[S(\mathbf{x})\phi(\mathbf{X})] = \alpha \mathrm{E}_{\theta_0}[S(\mathbf{X})].$$

Il est plus simple d'exprimer

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } \left[\frac{\sqrt{n}(\bar{x}-\theta_0)}{\sigma}\right]^2 > c \\ 0, & \text{sinon,} \end{cases}$$

car

$$\left[\frac{\sqrt{n}(\bar{x}-\theta_0)}{\sigma}\right]^2 > c$$

équivaut à

$$\bar{x} < \theta_0 - \frac{\sigma c}{\sqrt{n}} \text{ ou } \bar{x} > \theta_0 + \frac{\sigma c}{\sqrt{n}}$$

Il suffit alors de trouver c qui satisfait aux relations précédentes, ou, plus simplement tel que $\mathrm{P}(Y > c) = \alpha$, $Y \sim \chi^2(1)$. En effet sous (H_0), $\sqrt{n}(\bar{X} - \theta_0)/\sigma \sim \mathcal{N}(0, 1)$.

Remarque : En beaucoup de situations pratiques il y a un paramètre réel important θ mais aussi d'autres paramètres réels $\vartheta_1, \ldots, \vartheta_k$ moins intéressants (*paramètres de nuisance*). La densité est de type exponentiel :

$$f_{\theta,\vartheta_1,\ldots,\vartheta_k}(x) = h(x) \cdot \exp\left\{\theta U(x) + \vartheta_1 U_1(x) + \ldots + \vartheta_k U_k(x) + V(\theta, \vartheta_1, \ldots, \vartheta_k)\right\},$$

où $\theta \in \Theta \subset \mathbb{R}$. Les tests portent sur le paramètre θ, les autres paramètres restent non déterminés. Ainsi, si on fixe $\theta_0, \theta_1, \theta_2 \in \Theta$, $\theta_1 < \theta_2$ on peut tester

$$\left\{\begin{array}{l} (\mathrm{H}_1) : \theta \in \Theta_0 = \{\theta \in \Theta : \theta \leqslant \theta_0\} \\ (\mathrm{H}_1') : \theta \in \Theta_0 = \{\theta \in \Theta : \theta \geqslant \theta_0\} \\ (\mathrm{H}_2) : \theta \in \Theta_0 = \{\theta \in \Theta : \theta \leqslant \theta_1 \text{ ou } \theta \geqslant \theta_2\} \\ (\mathrm{H}_2') : \theta \in \Theta_0 = \{\theta \in \Theta : \theta_1 \leqslant \theta \leqslant \theta_2\} \\ (\mathrm{H}_0) : \theta \in \Theta_0 = \{\theta_0\} \end{array}\right\} \text{ contre } (\mathrm{A_i})((\mathrm{A_i'})) : \theta \in \Theta_0^c, i = 0, 1, 2.$$

Sous certaines hypothèses de régularité on peut construire des tests u.p.p.s.b. On donnera des exemples concernant les lois gaussiennes dans le paragraphe suivant.

2.3 Tester les paramètres des lois gaussiennes

Dans ce paragraphe la loi commune des $X_1, \ldots, X_n$ est gaussienne $\mathcal{N}(m, \sigma^2)$, où les deux paramètres m et σ^2 sont inconnus. Seulement un des deux paramètres sera important, l'autre peut être vu comme paramètre de nuisance. On mettra le densité sous la forme de la Remarque précédente et il est possible de prouver que les hypothèses (non précisées) de régularité sont remplies. Ainsi les tests qu'on va décrire seront des tests u.p.p.s.b. (fait admis). Tous les tests sont de niveau $\alpha \in]0, 1[$.

A. Tests pour la variance

i) On veut tester

$$(\mathrm{H}) : \sigma_1 \leqslant \sigma \leqslant \sigma_2 \text{ contre } (\mathrm{A}) : \sigma < \sigma_1 \text{ ou } \sigma > \sigma_2.$$

Le test est donné par

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } \sum_{j=1}^n (x_j - \bar{x})^2 < c_1 \text{ ou } \sum_{j=1}^n (x_j - \bar{x})^2 > c_2 \\ 0, & \text{sinon,} \end{cases}$$

où c_1, c_2 sont données par

$$\mathrm{P}(Y < \frac{c_1}{\sigma_i^2}) + \mathrm{P}(Y > \frac{c_2}{\sigma_i^2}) = \alpha, i = 1, 2,$$

où $Y = \frac{1}{\sigma^2} \sum_{j=1}^n (X_j - \bar{X})^2 \sim \chi^2(n-1)$. Par exemple, pour $n = 25$, $\sigma_1 = 2$, $\sigma_2 = 3$ et $\alpha = 0,05$, on peut écrire

$$\begin{cases} \mathrm{P}(Y < \frac{c_1}{4}) - \mathrm{P}(Y \leqslant \frac{c_2}{4}) = -0,95 \\ p(Y < \frac{c_1}{9}) - \mathrm{P}(Y \leqslant \frac{c_2}{9}) = -0,95 \end{cases}$$

et on cherche c_1, c_2 en utilisant la table de $Y \sim \chi^2(24)$ (par plusieurs essais).
Exercice : Suivant cet exemple, construire le test pour tester

$$(\mathrm{H}) : \sigma \leqslant \sigma_0 \text{ contre } (\mathrm{A}) : \sigma > \sigma_0$$

au niveau $\alpha = 0,05$. On donne $n = 25$, $\sigma_0 = 3$. Trouver la puissance pour $\sigma = 5$. Ce test est-il u.p.p. ? De la même façon construire les tests pour tester

$$(\mathrm{H}) : \sigma \geqslant \sigma_0 \text{ contre } (\mathrm{A}) : \sigma < \sigma_0$$

et pour tester

$$(\mathrm{H}) : \sigma \leqslant \sigma_1 \text{ ou } \sigma \geqslant \sigma_2 \text{ contre } (\mathrm{A}) : \sigma_1 < \sigma < \sigma_2.$$

ii) On veut tester

$$(\mathrm{H}) : \sigma = \sigma_0 \text{ contre } (\mathrm{A}) : \sigma \neq \sigma_0.$$

Le test est donné par

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } \sum_{j=1}^n (x_j - \bar{x})^2 < c_1 \text{ ou } \sum_{j=1}^n (x_j - \bar{x})^2 > c_2 \\ 0, & \text{sinon,} \end{cases}$$

où c_1, c_2 sont données par (exercice)

$$\int_{c_1/\sigma_0^2}^{c_2/\sigma_0^2} g(t)dt = \frac{1}{n-1} \int_{c_1/\sigma_0^2}^{c_2/\sigma_0^2} tg(t)dt = 1 - \alpha,$$

où g est la densité de $Y \sim \chi^2(n-1)$.

B. Tests pour l'espérance.

i) On veut tester

$$(\mathrm{H}) : m \leqslant m_0 \text{ contre } (\mathrm{A}) : m > m_0.$$

On construit le test

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } t(\mathbf{x}) > c \\ 0, & \text{sinon,} \end{cases}$$

où

$$t(\mathbf{x}) = \frac{\sqrt{n}(\bar{x} - m_0)}{\sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2}}$$

et où c est déterminée par

$$\mathrm{P}(T > c) = \alpha,\ T \sim \text{Student}(n-1).$$

Pour $n = 25$ et $\alpha = 0,05$ on a $\mathrm{P}(T > c) = 0,05$ avec $T \sim \text{Student}(24)$; donc $c = 1,7109$.
ii) On veut tester

$$(\mathrm{H}) : m = m_0 \text{ contre } (\mathrm{A}) : m \neq m_0.$$

On construit le test

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } t(\mathbf{x}) < -c \text{ ou } t(\mathbf{x}) > c,\ (c > 0) \\ 0, & \text{sinon,} \end{cases}$$

et où c est déterminée par

$$\mathrm{P}(T > c) = \frac{\alpha}{2},\ T \sim \text{Student}(n-1).$$

Si on veut déterminer la puissance on a besoin de la loi de Student décentrée (voir Définition 2.10, §2.6) :

$$T' = \frac{\xi}{\sqrt{\frac{1}{k}(\xi_1^2 + \ldots + \xi_k^2)}} \sim \text{Student}(k, m),$$

où $\xi, \xi_1, \ldots, \xi_k$ indépendantes de lois gaussiennes, $\xi \sim \mathcal{N}(m, 1)$, $\xi_1, \ldots, \xi_k \sim \mathcal{N}(0, 1)$.

REMARQUE : Considérons maintenant $X_1, \ldots, X_r$ un r-échantillon de loi $\mathcal{N}(m_1, \sigma_1^2)$ et $Y_1, \ldots, Y_n$ un n-échantillon de loi $\mathcal{N}(m_2, \sigma_2^2)$, les deux indépendants entre eux. Les quatre paramètres sont inconnus. Soit $m = m_1 - m_2$ et $\tau = \sigma_2^2/\sigma_1^2$. On veut construire des tests concernant les paramètres m et τ (il y aura un d'intérêt et l'autre de nuisance) ; les tests qu'on obtient seront u.p.p.s.b. Dans le paragraphe suivant on donne une méthode plus générale de construction et on retrouve les mêmes tests.

2.4 Méthode de construction des tests

Soit $\mathbf{X}$ un n-échantillon de densité $f_\theta(\mathbf{x})$, $\theta \in \Theta \subset \mathbb{R}^d$, et soit $\Theta_0 \subset \Theta$. On a vu que lorsque Θ_0 et Θ_0^c ont un seul point, disons respectivement θ_0 et θ_1 le rapport $R(\mathbf{x}) = f_{\theta_0}(\mathbf{x})/f_{\theta_1}(\mathbf{x})$, suffit pour construire un test. Ainsi pour tester (H) : $\theta \in \Theta_0$ contre (A) : $\theta \in \theta_0^c$ dans cette situation, le test p.p. rejette (H) lorsque le **rapport de vraisemblance** $R(\mathbf{x})$ est plus petit q'une constante c déterminée par le niveau du test. En effet, pour le cas discret, $f_{\theta_i}(\mathbf{x}^{\text{obs}})$ est la probabilité d'observer $\mathbf{x}^{\text{obs}}$ sous P_{θ_i}. Il est donc naturel de rejeter (H) si $f_{\theta_0}(\mathbf{x}) \leqslant f_{\theta_1}(\mathbf{x})$, ou encore si $f_{\theta_0}(\mathbf{x})/f_{\theta_1}(\mathbf{x}) \leqslant c \leqslant 1$

Il est clair que lorsque Θ_0 et Θ_0^c ont plus d'un point cette méthode ne s'applique plus. Toutefois on peut s'y ramener en utilisant l'idée suivante.

On va comparer cette fois $\sup\{f_\theta(\mathbf{x}) : \theta \in \Theta_0\}$ et $\sup\{f_\theta(\mathbf{x}) : \theta \in \Theta_0^c\}$. Il est donc raisonnable de rejeter (H) si l'observation $\mathbf{x}^{\text{obs}}$ est beaucoup plus probable quand le paramètre θ varie dans Θ_0^c que lorsqu'il varie dans Θ_0, c'est-à-dire, quand

$$\sup\{f_\theta(\mathbf{x}) : \theta \in \Theta_0^c\} >> \sup\{f_\theta(\mathbf{x}) : \theta \in \Theta_0\},$$

autrement dit quand le rapport

$$R(\mathbf{x}) = \frac{\sup\{f_\theta(\mathbf{x}) : \theta \in \Theta_0\}}{\sup\{f_\theta(\mathbf{x}) : \theta \in \Theta_0^c\}}$$

est très petit. Souvent on considère plutôt le **rapport de vraisemblance**

$$\check{R}(\mathbf{x}) := \frac{\sup\limits_{\theta \in \Theta_0} f_\theta(\mathbf{x})}{\sup\limits_{\theta \in \Theta} f_\theta(\mathbf{x})} \, (0 < \check{R} \leqslant 1)$$

et on rejette (H) quand le rapport de vraisemblance est trop petit, c'est à dire $\leqslant c$, où $0 \leqslant c \leqslant 1$ est donnée par le niveau du test. En effet on peut montrer que :

$$\check{R}(\mathbf{x}) = \min\{1, R(\mathbf{x})\}$$

d'où on peut montrer que $R(\mathbf{x}) \leqslant c$ si et seulement si $\check{R}(\mathbf{x}) \leqslant c$. Désormais on notera seulement par R cette dernière quantité.

Donnons l'interprétation intuitive du fait de rejeter (H) quand le rapport de vraisemblance $R(\mathbf{x})$ est petit. Considérons pour simplifier le cas discret. On a vu que la quantité $\sup\{f_\theta(\mathbf{x}) : \theta \in \Theta_0\}$ est le maximum de la probabilité d'observer $\mathbf{x}$, si θ varie dans Θ_0. De la même façon $\sup\{f_\theta(\mathbf{x}) : \theta \in \Theta\}$ est le maximum de la probabilité d'observer $\mathbf{x}$ sans aucune restriction. Ainsi, si $\theta \in \Theta_0$, ces deux maximums auront tendance à être proches l'un de l'autre (en supposant la continuité en θ de la vraisemblance $f_\theta(\mathbf{x})$), et donc R sera proche de 1. Lorsque R est trop loin de 1, l'observation a tendance à discréditer (H) et donc à la rejeter.

Définition 2.8 *Pour tester* (H) $\theta \in \Theta_0$ *contre* (A) $\theta \in \Theta_0^c$, *on construit la statistique de décision*

$$R(\mathbf{x}) := \frac{\sup\limits_{\theta\in\Theta_0} f_\theta(\mathbf{x})}{\sup\limits_{\theta\in\Theta} f_\theta(\mathbf{x})} \tag{12}$$

et le **test de rapport de vraisemblance** *a pour région de rejet* $\{\mathbf{x} \in \mathbb{R}^n : R(\mathbf{x}) \leqslant c\}$, *avec* $0 \leqslant c \leqslant 1$.

Le test donné par le lemme de Neyman-Pearson est un test de rapport de vraisemblance.

Dans la mise en place du test de rapport de vraisemblance on rencontre deux types de difficultés. La première est de calculer les deux maximums sur Θ_0 et Θ. Le calcul du maximum sur Θ revient à trouver l'estimateur du maximum de vraisemblance pour θ. Le calcul du maximum sur Θ_0 reste souvent un problème difficile, mais quand l'hypothèse (H) est simple le problème ne se pose pas. La deuxième difficulté est de trouver la constante c, donc de connaître la loi de $R(\mathbf{X})$. Il y a des cas particuliers où cette loi peut être trouvée. Dans d'autres cas on est seulement capables de dire la loi asymptotique (c'est-à-dire on peut calculer la limite en loi de $R(\mathbf{X})$ quand $n \to \infty$) donc de trouver une valeur approximative de c quand n est grand.

Malgré ces difficultés il s'agit quand même d'une méthode unifiée de construction de tests. Dans certaines situations pratiques les tests construits coïncident avec les tests dont on connaît les propriétés optimales (u.p.p. ou u.p.p.s.b.).

Enfin, notons qu'on peut aussi utiliser la statistique $-2\ln R(\mathbf{X})$ à la place de $R((\mathbf{X})$ (sous certaines hypothèses de régularité on connaît sa loi asymptotique, sous (H)). En termes de cette statistique, on rejette (H) lorsque $-2\ln R(\mathbf{x}) > c$, où c est donnée par le niveau du test.

Dans les exemples suivants on reprend le problème de **tester les paramètres des lois gaussiennes** :

A. Tests pour l'espérance.
Soit $X_1, \ldots, X_n$ un n-échantillon de loi $\mathcal{N}(m, \sigma^2)$.
i) On suppose σ connu et on veut tester l'hypothèse (H) : $m \in \Theta_0 = \{m_0\}$. Ici le paramètre est $m \in \Theta = \mathbb{R}$. On sait que l'estimateur du maximum de vraisemblance de m est $\hat{m} = \bar{X}$.

On a

$$\sup_{m\in\Theta} f_m(\mathbf{X}) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^n (X_j - \bar{X})^2\right], \ \Theta = \mathbb{R}$$

et

$$\sup_{m\in\Theta_0} f_m(\mathbf{X}) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^n (X_j - m_0)^2\right], \ \Theta_0 = \{m_0\}.$$

Il est plus facile de trouver la loi de $-2\ln R(\mathbf{X})$ plutôt que celle de $R(\mathbf{X})$. En fait on peut voir (exercice)

$$-2\ln R(\mathbf{X}) = \frac{n}{\sigma^2}(\bar{X} - m_0)^2$$

et le test du rapport de maximum de vraisemblance est équivalent à

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } \left[\frac{\sqrt{n}(\bar{x}-m_0)}{\sigma}\right]^2 > c \\ 0, & \text{sinon,} \end{cases}$$

où c est donnée par

$$\mathrm{P}(Y > c) = \alpha, \ Y \sim \chi^2(1).$$

Il s'agit du même test trouve dans l'exemple du paragraphe 2.3 donc il est u.p.p.s.b.
ii) Supposons cette fois-ci que σ est aussi inconnu. On veut tester la même hypothèse (H) : $m = m_0$ qui est maintenant composite car σ est inconnu. Précisément

$$\Theta = \{\theta = (m, \sigma^2) : m \in \mathbb{R}, \sigma^2 > 0\} \ \text{ et } \ \Theta_0 = \{\theta = (m, \sigma^2) : m = m_0, \sigma^2 > 0\}$$

d'où les estimateurs de maximum de vraisemblance suivants

$$\hat{\sigma}^2_{\Theta} = \frac{1}{n}\sum_{j=1}^n (X_j - \bar{X})^2, \ \hat{\sigma}^2_{\Theta_0} = \frac{1}{n}\sum_{j=1}^n (X_j - m_0)^2.$$

On trouve

$$\sup_{m\in\Theta} f_m(\mathbf{X}) = \frac{1}{\left(\sqrt{2\pi\hat{\sigma}^2_{\Theta}}\right)^n} \exp\left[-\frac{1}{2\hat{\sigma}^2_{\Theta}}\sum_{j=1}^n (X_j - \bar{X})^2\right] = \frac{1}{\left(\sqrt{2\pi\hat{\sigma}^2_{\Theta}}\right)^n} e^{-n/2}$$

et

$$\sup_{m\in\Theta_0} f_m(\mathbf{X}) = \frac{1}{\left(\sqrt{2\pi\hat{\sigma}^2_{\Theta_0}}\right)^n} \exp\left[-\frac{1}{2\hat{\sigma}^2_{\Theta_0}}\sum_{j=1}^n (X_j - m_0)^2\right] = \frac{1}{\left(\sqrt{2\pi\hat{\sigma}^2_{\Theta_0}}\right)^n} e^{-n/2}.$$

On trouve

$$R(\mathbf{X}) = \left(\frac{\hat{\sigma}^2_{\Theta}}{\hat{\sigma}^2_{\Theta_0}}\right)^{n/2}, \ \text{ ou } \ R(\mathbf{X})^{2/n} = \frac{\sum_{j=1}^n (X_j - \bar{X})^2}{\sum_{j=1}^n (X_j - m_0)^2}$$

et comme $\sum_{j=1}^n (X_j - m_0)^2 = \sum_{j=1}^n (X_j + \bar{X})^2 + n(\bar{X} - m_0)^2$, on trouve

$$R(\mathbf{X})^{2/n} = \left[1 + \frac{1}{n-1} \frac{n(\bar{X} - m_0)^2}{\frac{1}{n-1}\sum_{j=1}^n (X_j - \bar{X})^2}\right]^{-1} = \left(1 + \frac{t(\mathbf{X})^2}{n-1}\right)^{-1},$$

où

$$t(\mathbf{X}) = \frac{\sqrt{n}(\bar{X} - m_0)}{\sqrt{\frac{1}{n-1}\sum_{j=1}^n (X_j - \bar{X})^2}}.$$

La condition que le rapport de vraisemblance soit plus petit qu'une constante est équivalente à $t(\mathbf{x})^2 > c$ pour une certaine constante c. Alors le test du rapport de vraisemblance est équivalent au test

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } t(\mathbf{x}) < -c \text{ ou } t(\mathbf{x}) > c \\ 0, & \text{sinon,} \end{cases}$$

où c est donnée par $\mathrm{P}(T > c) = \alpha/2$ avec $T \sim \text{Student}(n-1)$. On retrouve le test à la fin du paragraphe 2.4 donc ce test est u.p.p.s.b.

B. Comparaison des paramètres de deux lois gaussiennes.

Soit $X_1, \ldots, X_r$ un r-échantillon de loi $\mathcal{N}(m_1, \sigma_1^2)$ indépendant d'un n-échantillon $Y_1, \ldots, Y_s$ de loi $\mathcal{N}(m_2, \sigma_2^2)$. La densité jointe de $(\mathbf{X}, \mathbf{Y})$ est donnée par

$$f_{m_1,m_2,\sigma_1^2,\sigma_2^2}(\mathbf{x}, \mathbf{y}) = \frac{1}{(\sqrt{2\pi})^{r+s}} \frac{1}{(\sigma_1^2)^r (\sigma_2^2)^s} \exp\left[-\frac{1}{2\sigma_1^2}\sum_{i=1}^r (x_i - m_1)^2 - \frac{1}{2\sigma_2^2}\sum_{j=1}^s (y_j - m_2)^2\right].$$

i) (**comparer deux espérances**)
On suppose que $\sigma_1 = \sigma_2 = \sigma$ inconnu et on veut tester l'hypothèse (H) : $m_1 = m_2 (= m$ non précisé). On pose

$$\Theta = \{\theta = (m_1, m_2, \sigma^2) : m_1, m_2 \in \mathbb{R}, \sigma^2 > 0\}$$

et

$$\hat{\theta}_\Theta := \left(\hat{m}_{1,\Theta} = \bar{X},\, \hat{m}_{1,\Theta} = \bar{Y},\, \hat{\sigma}^2_\Theta = \frac{1}{r+s}\left[\sum_{i=1}^r (X_i - \bar{X})^2 + \sum_{j=1}^s (Y_j - \bar{Y})^2\right]\right),$$

donc

$$\sup_{\theta\in\Theta} f_\theta(\mathbf{X}, \mathbf{X}) = \frac{1}{\left(\sqrt{2\pi\hat{\sigma}^2_\Theta}\right)^{r+s}} e^{-(r+s)/2}.$$

Par ailleurs

$$\Theta_0 = \{\theta = (m_1, m_2, \sigma^2) : m_1 = m_2 \in \mathbb{R}, \sigma^2 > 0\}$$

et les estimateurs de maximum de vraisemblance sont

$$\hat{m}_{\Theta_0} = \frac{1}{r+s}\left(\sum_{i=1}^{r} X_i + \sum_{j=1}^{s} Y_j\right) = \frac{r\bar{X} + s\bar{Y}}{r+s}$$

et

$$\hat{\sigma}^2_{\Theta_0} = \frac{1}{r+s}\left[\sum_{i=1}^{r}(X_i - \hat{m}_{\Theta_0})^2 + \sum_{j=1}^{s}(Y_j - \hat{m}_{\Theta_0})^2\right]$$

donc

$$\hat{\theta}_{\Theta_0} := \left(\hat{m}_{\Theta_0}, \hat{m}_{\Theta_0}, \hat{\sigma}^2_{\Theta_0}\right)$$

et

$$\sup_{\theta\in\Theta_0} f_\theta(\mathbf{X},\mathbf{Y}) = \frac{1}{\left(\sqrt{2\pi\hat{\sigma}^2_{\Theta_0}}\right)^{r+s}} e^{-(r+s)/2}.$$

On déduit

$$R(\mathbf{X},\mathbf{Y}) = \left(\frac{\hat{\sigma}^2_{\Theta}}{\hat{\sigma}^2_{\Theta_0}}\right)^{(r+s)/2}, \text{ ou } R(\mathbf{X},\mathbf{Y})^{2/(r+s)} = \frac{\hat{\sigma}^2_{\Theta}}{\hat{\sigma}^2_{\Theta_0}}.$$

Par des calculs simples

$$\sum_{i=1}^{r}(X_i - \hat{m}_{\Theta_0})^2 = \sum_{i=1}^{r}(X_i - \bar{X})^2 + \frac{rs^2}{(r+s)^2}(\bar{X} - \bar{Y})^2$$

et

$$\sum_{j=1}^{s}(Y_j - \hat{m}_{\Theta_0})^2 = \sum_{j=1}^{s}(Y_j - \bar{Y})^2 + \frac{r^2 s}{(r+s)^2}(\bar{X} - \bar{Y})^2$$

donc

$$(r+s)\hat{\sigma}^2_{\Theta_0} = (r+s)\hat{\sigma}^2_{\Theta} + \frac{rs}{r+s}(\bar{X}-\bar{Y})^2 = \sum_{i=1}^{r}(X_i - \bar{X})^2 + \sum_{j=1}^{s}(Y_j - \bar{Y})^2 + \frac{rs}{r+s}(\bar{X}-\bar{Y})^2.$$

On en déduit

$$R(\mathbf{X},\mathbf{Y})^{2/(r+s)} = \left(1 + \frac{t(\mathbf{X},\mathbf{Y})^2}{r+s-2}\right)^{-1},$$

où

$$t(\mathbf{X},\mathbf{Y}) = \frac{\sqrt{\frac{rs}{r+s}}(\bar{X}-\bar{Y})}{\sqrt{\frac{1}{r+s-2}\left[\sum_{i=1}^{r}(X_i - \bar{X})^2 + \sum_{j=1}^{s}(Y_j - \bar{Y})^2\right]}}$$

Le test du rapport de vraisemblance selon lequel on rejette (H) lorsque $R(\mathbf{x},\mathbf{y})$ est plus petit qu'une constante est équivalent au test

$$\phi(\mathbf{x},\mathbf{y}) = \begin{cases} 1, & \text{si } t(\mathbf{x}) < -c \text{ ou } t(\mathbf{x}) > c \\ 0, & \text{sinon,} \end{cases}$$

où c est donnée par $\mathrm{P}(T > c) = \alpha/2$ avec $T \sim \text{Student}(r+s-2)$. On peut montrer que ce test est u.p.p.s.b.

ii) (**comparer deux variances**)
On veut tester l'hypothèse (H) : $\sigma_1^2 = \sigma_2^2 (= \sigma^2$ non précisé). On pose

$$\Theta = \{\theta = (m_1, m_2, \sigma_1^2, \sigma_2^2) : m_1, m_2 \in \mathbb{R}, \sigma_1^2, \sigma_2^2 > 0\}$$

et on a les estimateurs de maximum de vraisemblance

$$\hat{m}_{1,\Theta} = \bar{X},\ \hat{m}_{2,\Theta} = \bar{Y},$$

$$\hat{\sigma}_{1,\Theta}^2 = \frac{1}{r}\sum_{i=1}^{r}(X_i - \bar{X})^2,\ \hat{\sigma}_{2,\Theta}^2 = \frac{1}{s}\sum_{j=1}^{n}(Y_j - \bar{Y})^2$$

tandis que

$$\Theta_0 = \{\theta = (m_1, m_2, \sigma^2) : m_1, m_2 \in \mathbb{R}, \sigma_1^2 = \sigma_2^2 > 0\}$$

et on a les estimateurs de maximum de vraisemblance

$$\hat{m}_{1,\Theta_0} = \bar{X},\ \hat{m}_{2,\Theta_0} = \bar{Y},$$

et

$$\hat{\sigma}_{\Theta_0}^2 = \frac{1}{r+s}\left[\sum_{i=1}^{r}(X_i - \bar{X})^2 + \sum_{j=1}^{s}(Y_j - \bar{Y})^2\right],$$

On déduit

$$\sup_{\theta\in\Theta} f_\theta(\mathbf{X}, \mathbf{Y}) = \frac{1}{(\sqrt{2\pi})^{r+s}} \frac{1}{\left(\hat{\sigma}_{1,\Theta}^2\right)^{r/2}\left(\hat{\sigma}_{2,\Theta}^2\right)^{s/2}} e^{-(r+s)/2}$$

et

$$\sup_{\theta\in\Theta_0} f_\theta(\mathbf{X}, \mathbf{Y}) = \frac{1}{(\sqrt{2\pi})^{r+s}} \frac{1}{\left(\hat{\sigma}_{\Theta_0}^2\right)^{(r+s)/2}} e^{-(r+s)/2}.$$

Enfin

$$R(\mathbf{X}, \mathbf{Y}) = \frac{\left(\hat{\sigma}_{1,\Theta}^2\right)^{r/2}\left(\hat{\sigma}_{2,\Theta}^2\right)^{s/2}}{\left(\hat{\sigma}_{\Theta_0}^2\right)^{(r+s)/2}}$$

$$= \frac{(r+s)^{(r+s)/2}\left[\sum_{i=1}^{r}(X_i - \bar{X})^2 / \sum_{j=1}^{s}(Y_j - \bar{Y})^2\right]^{r/2}}{r^{r/2}s^{s/2}\left\{\left[\sum_{i=1}^{r}(X_i - \bar{X})^2 + \sum_{j=1}^{s}(Y_j - \bar{Y})^2\right] / \sum_{j=1}^{s}(Y_j - \bar{Y})^2\right\}^{(r+s)/2}}$$

$$= \frac{(r+s)^{(r+s)/2}}{r^{r/2}s^{s/2}} \cdot \frac{\left[\frac{r-1}{s-1} \cdot \frac{\sum_{i=1}^{r}(X_i - \bar{X})^2/(r-1)}{\sum_{j=1}^{s}(Y_j - \bar{Y})^2/(s-1)}\right]^{r/2}}{\left[1 + \frac{r-1}{s-1} \cdot \frac{\sum_{i=1}^{r}(X_i - \bar{X})^2/(r-1)}{\sum_{j=1}^{s}(Y_j - \bar{Y})^2/(s-1)}\right]^{(r+s)/2}}$$

$$= \frac{(r+s)^{(r+s)/2}}{r^{r/2}s^{s/2}} \cdot \frac{\left(\frac{r-1}{s-1}f(\mathbf{X},\mathbf{Y})\right)^{r/2}}{\left[1+\frac{r-1}{s-1}f(\mathbf{X},\mathbf{Y})\right)^{(r+s)/2}},$$

où

$$f(\mathbf{X},\mathbf{Y}) = \frac{\sum_{i=1}^{r}(X_i-\bar{X})^2/(r-1)}{\sum_{j=1}^{s}(Y_j-\bar{Y})^2/(s-1)}$$

Le test du maximum de vraisemblance est équivalent au test basé sur f et on rejette (H) lorsque

$$\frac{\left(\frac{r-1}{s-1}f(\mathbf{X},\mathbf{Y})\right)^{r/2}}{\left(1+\frac{r-1}{s-1}f(\mathbf{X},\mathbf{Y})\right)^{(r+s)/2}} < c \text{ pour une certaine constante } c.$$

Notons

$$g(f) = \frac{\left(\frac{r-1}{s-1}f\right)^{r/2}}{\left(1+\frac{r-1}{s-1}f\right)^{(r+s)/2}}$$

et on a $g(0)=0$, $\lim_{f\to\infty} g(f)=0$. Cette fonction a un maximum au point $f_{\max} = \frac{r(s-1)}{s(r-1)}$, est croissante entre 0 et $f_{\max}$ et décroissante entre $f_{\max}$ et ∞. Ainsi

$$g(f) < c \text{ si et seulement si } f < c_1 \text{ ou } f > c_2,$$

pour certaines constantes c_1, c_2.

On peut voir aussi que, sous (H), $f(\mathbf{X},\mathbf{Y}) \sim \mathcal{F}(r-1,s-1)$. On trouve les constantes c_1, c_2 à l'aide des conditions

$$\mathrm{P}(F<c_1) = \mathrm{P}(F>c_2) = \frac{\alpha}{2},\ F \sim \mathcal{F}(r-1,s-1).$$

2.5 Tests et intervalles de confiance

Il y a une manière de construire des intervalles de confiance à l'aide des tests :

Théorème 2.6 *Soit* $\mathbf{X}$ *un* n*-échantillon de loi* $\{\mathrm{P}_\theta; \theta \in \Theta\}$*. Pour chaque* $\theta_0 \in \Theta$ *on considère le problème de tester* (H) : $\theta = \theta_0$ *au seuil* α *et soit* $\mathcal{A}(\theta_0)$ *sa région d'acceptation. Pour chaque* $\mathbf{x} \in \mathbb{R}^n$ *on définit* $\mathcal{C}(\mathbf{x}) \subset \Theta$ *par*

$$\mathcal{C}(\mathbf{x}) := \{\theta_0 \in \Theta : \mathbf{x} \in \mathcal{A}(\theta_0)\}. \tag{13}$$

Alors $\mathcal{C}(\mathbf{X})$ *est une région de confiance pour* θ *de seuil de sécurité* $1-\alpha$*. Réciproquement, soit* $\mathcal{C}(\mathbf{X})$ *une région de confiance pour* θ *de seuil de sécurité* $1-\alpha$*. Pour tout* $\theta_0 \in \Theta$ *on définit*

$$\mathcal{A}(\theta_0) := \{\mathbf{x} : \theta_0 \in \mathcal{C}(\mathbf{x})\}. \tag{14}$$

Alors $\mathcal{A}(\theta_0)$ *est la région d'acceptation d'un tests de seuil* α *pour* (H) : $\theta = \theta_0$.

Preuve. Pour la partie directe : comme $\mathcal{A}(\theta_0)$ est la région d'acceptation d'un test de seuil α on a

$$\mathrm{P}_{\theta_0}\left(\mathbf{X} \notin \mathcal{A}(\theta_0)\right) \leqslant \alpha \Rightarrow \mathrm{P}_{\theta_0}\left(\mathbf{X} \in \mathcal{A}(\theta_0)\right) \geqslant 1-\alpha.$$

Comme θ_0 est quelconque, on écrit θ au lieu de θ_0. Alors, de l'inégalité précédente et de (13) on trouve que

$$\mathrm{P}_{\theta}\left(\theta \in \mathcal{C}(\mathbf{X})\right) = \mathrm{P}_{\theta}\left(\mathbf{X} \in \mathcal{A}(\theta)\right) \geqslant 1-\alpha,$$

donc $\mathcal{C}(\mathbf{X})$ est une région de confiance pour θ de seuil de sécurité $1-\alpha$.

Pour l'autre partie on voit que l'erreur de 1ère espèce pour (H) : $\theta = \theta_0$ avec la région d'acceptation $\mathcal{A}(\theta_0)$ est

$$\mathrm{P}_{\theta_0}\left(\mathbf{X} \notin \mathcal{A}(\theta_0)\right) = \mathrm{P}_{\theta_0}\left(\theta_0 \notin \mathcal{C}(\mathbf{X})\right) \leqslant \alpha,$$

donc le test est de seuil α pour $\theta = \theta_0$. □

2.6 Modèle linéaire gaussien

Définition 2.9 *Soit* $\mathbf{X} = (X_1, \ldots, X_n)^*$ *le vecteur des observations. On dira que* $\mathbf{X}$ *satisfait à un* **modèle linéaire gaussien** *si*

$$\mathbf{X} = A\theta + \varepsilon,$$

où $\theta \in \mathbb{R}^d$ *est le paramètre du modèle (*$d < n$ *pour qu'on ait au moins autant d'observations que de paramètres dans le modèle),* A *est une matrice* $n \times d$ *connue et* $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^*$ *est un vecteur gaussien* $\mathcal{N}_n(0, \sigma^2 I_n)$.

On suppose que σ^2 est inconnu ; le paramètre du modèle est $(\theta, \sigma^2) \in \mathbb{R}^d \times \mathbb{R}_+$. Si on note $\mathbf{m} = A\theta$, $\mathbf{m}$ varie dans le sous-espace vectoriel $V = \mathrm{Im}A$ de $\mathbb{R}^n$ avec $\mathrm{dim}V \leqslant p < n$. De plus pour tout $j \in \{1, \ldots, n\}$,

$$X_j = m_j + \varepsilon_j \sim \mathcal{N}(m_j, \sigma^2),$$

les variables X_i sont indépendantes et $\mathrm{E}\left((X_i - m_i)(X_j - m_j)\right) = (\sigma^2 I_n)_{ij} = \sigma^2 \delta_{ij}$.

Exemples :
1) n**-échantillon gaussien :** Soit $\mathbf{X}$ un n-échantillon de loi $\mathcal{N}(m, \sigma^2)$. Alors $X_j = m + \varepsilon_j$, avec $(\varepsilon_1, \ldots, \varepsilon_n)$ un n-échantillon de loi $\mathcal{N}(0, \sigma^2)$. Ainsi

$$\mathbf{X} = A\theta + \varepsilon,$$

avec $A = (1, \ldots, 1)^* = \mathbf{1}$.
2) régression linéaire : On suppose que pour

$$X_j = \beta + \gamma t_j + \varepsilon_j,\ j \in \{1, \ldots, n\}.$$

où $(\varepsilon_1, \ldots, \varepsilon_n)$ est n-échantillon de loi $\mathcal{N}(0, \sigma^2)$. Ici le paramètre est $\theta = (\beta, \gamma)^*$ et $\mathbf{t} = (t_1, \ldots, t_n)^*$ est un vecteur connu. Ainsi

$$\mathbf{X} = A\theta + \varepsilon,$$

avec

$$A = \begin{pmatrix} 1 & t_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & t_n \end{pmatrix} = (\mathbf{1}\,\mathbf{t}) .$$

3) analyse de variance à un facteur : On observe k échantillons indépendants, de lois $\mathcal{N}(m_1, \sigma^2)$, …, $\mathcal{N}(m_k, \sigma^2)$; la taille du i-ième échantillon est n_i. On a donc pour tout $i \in \{1, \ldots, k\}$, pour tout $j \in \{1, \ldots, n_i\}$

$$X_{ij} = m_i + \varepsilon_{ij},$$

où (ε_{ij}) est un n-échantillon de loi $\mathcal{N}(0, \sigma^2)$ $(n = n_1 + \ldots + n_k)$. On a donc

$$\mathbf{X} = A\theta + \varepsilon,$$

où

$$\mathbf{X} = \begin{pmatrix} X_{11} \\ \cdot \\ \cdot \\ X_{1n_1} \\ \cdot \\ \cdot \\ X_{k1} \\ \cdot \\ \cdot \\ X_{kn_k} \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_{11} \\ \cdot \\ \cdot \\ \varepsilon_{1n_1} \\ \cdot \\ \cdot \\ \varepsilon_{k1} \\ \cdot \\ \cdot \\ \varepsilon_{kn_k} \end{pmatrix}, \theta = \begin{pmatrix} m_1 \\ \cdot \\ \cdot \\ m_k \end{pmatrix}, A = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ 1 & 0 & \ldots & 0 \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ 0 & 0 & \ldots & 1 \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ 0 & 0 & \ldots & 1 \end{pmatrix} \begin{array}{l} \left.\begin{matrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{matrix}\right\} n_1 \\ \cdot \\ \cdot \\ \left.\begin{matrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{matrix}\right\} n_k \end{array} .$$

On supposera A injective $\dim V = \mathrm{Im} A = d < n$. Pour estimer θ il suffit d'estimer $\mathbf{m} = A\theta \in V$ dans le modèle

$$\mathbf{X} = \mathbf{m} + \varepsilon,\ \mathbf{m} \in V,\ \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n).$$

Proposition 2.2 *L'estimateur du maximum de vraisemblance de (m, σ^2) dans le modèle précédent est :*

$$\hat{\mathbf{m}} = \Pi_V(\mathbf{X}),\ \hat{\sigma}^2 = \frac{1}{n}\|\mathbf{X} - \Pi_V(\mathbf{X})\|^2$$

où Π_V désigne la projection orthogonale sur V.

Preuve. La log-vraisemblance est

$$\ln f(\mathbf{x}, m, \sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{n}(X_j - m_j)^2.$$

Maximiser cette fonction m revient à minimiser $\sum_{j=1}^{n}(X_j - m_j)^2 = \|\mathbf{X} - \mathbf{m}\|^2$ sous la contrainte $\mathbf{m} \in V$. La solution est donnée par la méthode des moindres carrées $\hat{\mathbf{m}} = \Pi_V(\mathbf{X})$. Pour maximiser en σ^2, l'équation obtenue par dérivation est

$$\frac{n}{2\hat{\sigma}^2} - \frac{1}{2\hat{\sigma}^4}\|\mathbf{X} - \hat{\mathbf{m}}\|^2 = 0,$$

soit $\hat{\sigma}^2 = \frac{1}{n}\|\mathbf{X} - \Pi_V(\mathbf{X})\|^2$. □

Proposition 2.3 *Si A est injective, l'estimateur du maximum de vraisemblance pour (θ, σ^2) dans le modèle linéaire gaussien est donné par*

$$\hat{\theta} = (A^*A)^{-1}A^*\mathbf{X},\ \hat{\sigma}^2 = \frac{1}{n}\|\mathbf{X} - A(A^*A)^{-1}A^*\mathbf{X}\|^2.$$

Preuve. On pourra vérifier que A est injective si et seulement si A^*A est inversible. La log-vraisemblance est

$$\ln f(\mathbf{x}, \theta, \sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{j=1}^{n}(X_j - (A\theta)_j)^2.$$

Notons $\hat{\theta}$ et $\hat{\sigma}^2$ des estimateurs du maximum de vraisemblance du modèle. Alors $\hat{\mathbf{m}} = A\hat{\theta}$ minimise la fonction $\mathbf{m} \in \mathrm{Im}A \mapsto \|\mathbf{X} - \mathbf{m}\|^2$, c'est-à-dire $\hat{\mathbf{m}} = \Pi_V(\mathbf{X})$. De même que précédement, on a alors $\hat{\sigma}^2 = \frac{1}{n}\|\mathbf{X} - \hat{\mathbf{m}}\|^2$. Pour trouver $\hat{\theta}$ il suffit donc de résoudre l'équation (en $\hat{\theta}$) $A\hat{\theta} = \hat{\mathbf{m}} = \Pi_V(\mathbf{X})$ qui admet une solution unique puisque $\Pi_V(\mathbf{X}) \in \mathrm{Im}A$ et A est injective. On va calculer $\Pi_V(\mathbf{X})$. La décomposition de $\mathbf{x}$ sur $V \oplus V^\perp$ est $\mathbf{x} = A\mathbf{z} + \mathbf{w}$. Comme $V^\perp = \mathrm{Im}A^\perp = \mathrm{Ker}A^*$. On a

$$A^*\mathbf{x} = (A^*A)\mathbf{z} + A^*\mathbf{w} = (A^*A)\mathbf{z},\ \text{ d'où }\ \mathbf{z} = (A^*A)^{-1}A^*\mathbf{x}.$$

Ainsi $\Pi_V(\mathbf{x}) = A\mathbf{z} = A(A^*A)^{-1}A^*\mathbf{x}$. Alors $A\hat{\theta} = \Pi_V(\mathbf{X}) = A(A^*A)^{-1}A^*\mathbf{X}$. Alors $\hat{\theta} = (A^*A)^{-1}A^*\mathbf{X}$. □

Exemple : régresion linéaire. Par calcul

$$A^*A = n\begin{pmatrix} 1 & \bar{t} \\ \bar{t} & \frac{1}{n}\sum_{j=1}^{n} t_j^2 \end{pmatrix},$$

où $\bar{t} = \frac{1}{n}\sum_{j=1}^n t_j$. Ainsi

$$(A^*A)^{-1} = \frac{1}{n\mathrm{Var}(t)} \begin{pmatrix} \frac{1}{n}\sum_{j=1}^n t_j^2 & -\bar{t} \\ -\bar{t} & 1 \end{pmatrix},$$

où $\mathrm{Var}(t) = \frac{1}{n}\sum_{j=1}^n t_j^2 - \bar{t}^2 = \frac{1}{n}\sum_{j=1}^n (t_j - \bar{t})^2$. Alors

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} \bar{X}\frac{\frac{1}{n}\sum_{j=1}^n t_j^2}{\mathrm{Var}(t)} - \bar{t}\,\frac{\frac{1}{n}\sum_{j=1}^n t_j X_j}{\mathrm{Var}(t)} \\ \\ \frac{\mathrm{Cov}(t,X)}{\mathrm{Var}(t)} \end{pmatrix},$$

où $\mathrm{Cov}(t,X) = \frac{1}{n}\sum_{j=1}^n t_j X_j - \bar{t}\bar{X} = \frac{1}{n}\sum_{j=1}^n (t_j - \bar{t})(X_j - \bar{X})$. Donc

$$\hat{\gamma} = \frac{\mathrm{Cov}(t,X)}{\mathrm{Var}(t)},\ \hat{\beta} = \bar{X} - \hat{\gamma}\bar{t},\ \hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^n (X_j - \hat{\beta} - \hat{\gamma}t_j)^2.$$

Exemple : analyse de variance à un facteur On a

$$A^*A = \begin{pmatrix} n_1 & 0 & \dots & 0 \\ 0 & n_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & n_k \end{pmatrix},\ A^*\mathbf{X} = \begin{pmatrix} \sum_{j=1}^{n_1} X_{1j} \\ \sum_{j=1}^{n_2} X_{2j} \\ \cdot \\ \sum_{j=1}^{n_k} X_{kj} \end{pmatrix}$$

d'où

$$\begin{pmatrix} \hat{m}_1 \\ \cdot \\ \hat{m}_k \end{pmatrix} = \begin{pmatrix} X_{1,\bullet} \\ \cdot \\ X_{k,\bullet} \end{pmatrix},\ \text{ où } X_{i,\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$$

et

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i,j}(X_{ij} - X_{i,\bullet})^2.$$

Théorème 2.7 *(de Cochran)*
Soit G un vecteur gaussien $G \sim \mathcal{N}_n(0, \sigma^2 I_n)$. Soit une décomposition orthogonale de $\mathbb{R}^n = V_1 \oplus \ldots \oplus V_k$, $dimV_i = n_i$, $i = 1, \ldots, k$. Alors les vecteurs $\Pi_{V_1}(G), \ldots, \Pi_{V_k}(G)$ sont gaussiens indépendents et

$$\frac{\|\Pi_{V_i}(G)\|^2}{\sigma^2} \sim \chi^2(n_i),\ i = 1, \ldots, k.$$

Preuve. Soit (e_{ij}) une base orthonormée de $\mathbb{R}^n$ adaptée à la décomposition $\mathbb{R}^n = V_1 \oplus \ldots \oplus V_k$, c'est-à-dire $(e_{ij} : j = 1, \ldots, n_i)$ est une base orthonormée de V_i. Soit $G = \sum_{i,j} (G, e_{ij}) e_{ij}$ la décomposition de G dans cette nouvelle base. On pose $\eta_{ij} = (G, e_{ij})/\sigma$ et on obtient un vecteur gaussien $\eta = (\eta_{ij})_{i,j} \sim \mathcal{N}_n(0, I_n)$. Mais

$$\Pi_{V_i}(G) = \sum_{j=1}^{n_i} (G, e_{ij}) e_{ij} = \sigma \sum_{j=1}^{n_i} \eta_{ij} e_{ij}$$

ne dépend que des η_{ij} pour $j \in \{1, \ldots, n_i\}$. On en déduit que les vecteurs $\Pi_{V_i}(G)$ sont indépendants et que

$$\frac{\|\Pi_{V_i}(G)\|^2}{\sigma^2} = \sum_{j=1}^{n_i} \eta_{ij}^2$$

suit la loi $\chi^2(n_i)$. $\square$

On va construire des tests pour des **hypothèses linéaires** :

$$(\mathrm{H}) : \mathbf{m} \in W \;\text{ contre }\; (\mathrm{A}) : \mathbf{m} \notin W,$$

où $W \subset V$ est un sous-espace vectoriel. Ce genre de test permet de diminuer le nombre de paramètres du modèle linéaire gaussien.

Exemples :
1) régression linéaire La moyenne est supposée appartenir au sous-espace vectoriel

$$V = \mathrm{Vect}\left\{ \begin{pmatrix} 1 \\ \cdot \\ \cdot \\ 1 \end{pmatrix}, \begin{pmatrix} t_1 \\ \cdot \\ \cdot \\ t_n \end{pmatrix} \right\}.$$

Le test

$$(\mathrm{H}) : \beta = 0 \;\text{ contre }\; (\mathrm{A}) : \beta \neq 0$$

est un test d'hypothèse linéaire avec

$$W = \mathrm{Vect}\left\{ \begin{pmatrix} t_1 \\ \cdot \\ \cdot \\ t_n \end{pmatrix} \right\}.$$

2) analyse de variance à un facteur La moyenne $\mathbf{m}$ est supposée appartenir à

$$V = \mathrm{Vect}\left\{\begin{pmatrix}1\\ .\\ .\\ 1\\ .\\ .\\ 0\\ .\\ .\\ 0\end{pmatrix}, \ldots, \begin{pmatrix}0\\ .\\ .\\ 0\\ .\\ .\\ 1\\ .\\ .\\ 1\end{pmatrix}\right\}.$$

Le test

$$(\mathrm{H}) : m_1 = \ldots = m_k \ \text{ contre } \ (\mathrm{A}) : \exists i, j : m_i \neq m_j$$

est un test d'hypothèse linéaire avec

$$W = \mathrm{Vect}\left\{\begin{pmatrix}1\\ .\\ .\\ 1\end{pmatrix}\right\}.$$

On va construire un test de rapport de vraisemblance. Ici

$$\Theta = \{(\mathbf{m}, \sigma^2) \in V \times \mathbb{R}_+^*\},\ \Theta_0 = \{(\mathbf{m}, \sigma^2) \in W \times \mathbb{R}_+^*\}.$$

et la vraisemblance du modèle est

$$f_{\mathbf{m},\sigma^2}(\mathbf{x}) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x} - \mathbf{m}\|^2\right).$$

On a les estimateurs de maximum de vraisemblance

$$\hat{m}_\Theta = \Pi_V(\mathbf{X}),\ \hat{\sigma}^2_\Theta = \frac{1}{n}\|\mathbf{X} - \Pi_V(\mathbf{X})\|^2,\ \hat{m}_{\Theta_0} = \Pi_W(\mathbf{X}),\ \hat{\sigma}^2_{\Theta_0} = \frac{1}{n}\|\mathbf{X} - \Pi_W(\mathbf{X})\|^2$$

donc

$$\sup_{(\mathbf{m},\sigma^2)\in\Theta} f_{\mathbf{m},\sigma^2}(\mathbf{X}) = \frac{1}{\left(\sqrt{2\pi\hat{\sigma}^2_\Theta}\right)^n} e^{-n/2},\quad \sup_{(\mathbf{m},\sigma^2)\in\Theta_0} f_{\mathbf{m},\sigma^2}(\mathbf{X}) = \frac{1}{\left(\sqrt{2\pi\hat{\sigma}^2_{\Theta_0}}\right)^n} e^{-n/2}.$$

On trouve

$$R(\mathbf{X}) = \left(\frac{\|\mathbf{X} - \Pi_V(\mathbf{X})\|^2}{\|\mathbf{X} - \Pi_W(\mathbf{X})\|^2}\right)^{n/2} \ \text{ ou } R(\mathbf{X})^{2/n} = \frac{\|\mathbf{X} - \Pi_V(\mathbf{X})\|^2}{\|\mathbf{X} - \Pi_W(\mathbf{X})\|^2}.$$

Mais

$$\|\mathbf{X} - \Pi_W(\mathbf{X})\|^2 = \|\mathbf{X} - \Pi_V(\mathbf{X})\|^2 + \|\Pi_V(\mathbf{X}) - \Pi_W(\mathbf{X})\|^2,$$

donc

$$R(\mathbf{X})^{2/n} = \left(1 + \frac{\|\Pi_V(\mathbf{X}) - \Pi_W(\mathbf{X})\|^2}{\|\mathbf{X} - \Pi_V(\mathbf{X})\|^2}\right)^{-1} = (1 + f(\mathbf{X}))^{-1}$$

Le test du rapport de vraisemblance qui rejette (H) quand R est plus petit qu'un constante est équivalent au test

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } f(\mathbf{x}) > c \\ 0, & \text{sinon,} \end{cases}$$

où c est donnée par $\mathrm{P}_{(\mathrm{H})}(f(\mathbf{X}) > c) = \alpha$. On donc besoin de la loi de $f(\mathbf{X})$ sous (H), c'est-à-dire quand $\mathbf{m} \in W$. Mais si $\mathbf{m} \in W$

$$\Pi_V(\mathbf{X}) - \Pi_W(\mathbf{X}) = \Pi_V(\mathbf{X} - \mathbf{m}) - \Pi_W(\mathbf{X} - \mathbf{m}) = \Pi_{V \setminus W}(\mathbf{X} - \mathbf{m})$$

et

$$\mathbf{X} - \Pi_V(\mathbf{X}) = (\mathbf{X} - \mathbf{m}) - \Pi_V(\mathbf{X} - \mathbf{m}) = \Pi_{V^\perp}(\mathbf{X} - \mathbf{m})$$

Comme $\mathbf{X} - \mathbf{m} \sim \mathcal{N}_n(0, \sigma^2 I_n)$, par le théorème de Cochran $\Pi_V(\mathbf{X}) - \Pi_W(\mathbf{X})$ et $\mathbf{X} - \Pi_V(\mathbf{X})$ sont indépendants et

$$\frac{\|\Pi_V(\mathbf{X}) - \Pi_W(\mathbf{X})\|^2}{\sigma^2} \sim \chi^2(\mathrm{dim}V - \mathrm{dim}W), \quad \frac{\|\mathbf{X} - \Pi_V(\mathbf{X})\|^2}{\sigma^2} \sim \chi^2(n - \mathrm{dim}V).$$

On déduit que

$$\frac{n - \mathrm{dim}V}{\mathrm{dim}V - \mathrm{dim}W} f(\mathbf{X}) \sim \mathcal{F}(\mathrm{dim}V - \mathrm{dim}W, n - \mathrm{dim}V).$$

Ainsi c est donné par

$$\mathrm{P}\left(\frac{n - \mathrm{dim}V}{\mathrm{dim}V - \mathrm{dim}W} F > c\right) = \alpha, \; F \sim \mathcal{F}(\mathrm{dim}V - \mathrm{dim}W, n - \mathrm{dim}V).$$

Exercice : Appliquer ce résultat aux deux exemples, la régression linéaire et analyse de variance à un facteur.

Définition 2.10 *Pour étudier la fonction puissance on a besoin de la* **loi de chi-deux décentrée** $\chi^2(k, \lambda)$, *c'est-à-dire la loi de* $\xi_1^2 + \ldots + \xi_k^2$ *où les variables* ξ_i *sont indépendantes de lois* $\mathcal{N}(m_i, 1)$. *Le nombre de degrés de liberté est* k *et le paramètre de décentrage est* $\lambda = m_1^2 + \ldots + m_k^2$. *Aussi on utilise une* **loi de Fisher décentrée** *car le numérateur de* $f(\mathbf{X})$ *suit une loi* $\chi^2(\mathit{dim}V - \mathit{dim}W, \|\mathbf{m} - \Pi_W(\mathbf{m})\|^2/\sigma^2)$. *De même on peut introduire la* **loi de Student décentrée** *(voir §2.3, Exemple B).*

Remarque : Tout ce que nous avons fait repose sur l'hypothèse que ε est un n-échantillon de loi gaussienne $\mathcal{N}(0, \sigma^2)$, c'est-à-dire que les variables X_j ont toutes la même variance σ^2. Cette hypothèse n'est pas évidente à priori donc il faut commencer en fait par vérifier cette hypothèse donc de tester l'égalité des variances pour deux échantillons gaussiens (voir le paragraphe précédent).

Remarque : On utilise le modèle d'analyse de la variance à un facteur quand on étudie une seule variable d'intérêt (par exemple la quantité de maïs produite) en fonction d'un seul facteur (par exemple la qualité du sol). Mais si on veut prendre en compte plusieurs facteurs (par exemple aussi l'exposition au soleil, la quantité d'eau, ...) on a besoin d'un modèle plus compliqué. Dans **le modèle d'analyse de variance à deux facteurs** on donne $I \times J$ échantillons indépendants

$$\{X_{ijk} : i = 1, \dots, I,\ j = 1, \dots, J,\ k = 1, \dots, n_{ij}\}$$

de variables gaussiennes de moyennes m_{ij} et de même variance σ^2. Par exemple X_{ijk} est la quantité de maïs produite sur la k-ième parcelle de qualité i exposée en direction j :

$$X_{ijk} = m_{ij} + \varepsilon_{ijk}$$

où $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$, $n = \sum_{i,j} n_{ij}$ est le nombre d'observations et $p = IJ + 1$ est le nombre de paramètres. On peut disposer d'une seule observation par case, c'est-à-dire $n_{ij} = 1$ auquel cas le nombre de paramètres est supérieur au nombre d'observations et il faut faire des hypothèse sur les paramètres. Souvent on prend le *modèle additif* $m_{ij} = m + \beta_i + \gamma_j$ (on néglige l'interaction éventuelle entre les deux facteurs). Lorsqu'on a plus d'observations (*modèle équilibré* $n_{ij} = K \geqslant 2$) on prend

$$X_{ijk} = m_{ij} + \varepsilon_{ijk},\ m_{ij} = m + \alpha_i + \beta_j + \gamma_{ij},\ i = 1, \dots, I,\ j = 1, \dots, J,\ k = 1, \dots, K.$$

On peut reprendre les mêmes problèmes (estimation et tests) pour ce modèle.

2.7 Tests non paramétriques d'ajustement

Une étape importante dans l'analyse statistique est le choix du modèle. Mais comment choisir le modèle (la loi qui modélise un phénomène) ? La problématique des tests d'ajustement est la suivante. On donne un n-échantillon de loi inconnue Q et on désire vérifier si cette loi coïncide à une loi connue Q_0 (donnée). Si la loi fait partie d'une famille paramétrique on reprend les techniques déjà utilisées. Mais quand on dispose de peu d'informations sur la loi que faire ? On donnera ici deux méthodes permettant de répondre à cette question.

2.7.1 Test du χ^2

On observe une variable aléatoire discrète X susceptible de prendre r valeurs $a_1, \dots, a_r \in \mathbb{R}^d$ (le support de la loi est donc supposé fini et connu). On notera $\mathbf{p} = (p_1, \dots, p_r)$ le vecteur des probabilités de chaque valeur possible :

$$p_j = \mathrm{P}(X = a_j) = Q(\{a_j\}) \in]0, 1[, \; j = 1, \ldots, r.$$

Soit par ailleurs une probabilité $Q_0 = \sum_{j=1}^r \pi_j \delta_{a_j}$ de même support mais avec les probabilités $\pi = (\pi_1, \ldots, \pi_r)$ connues. La question est $Q = Q_0$? En d'autres termes $\mathbf{p} = \pi$ vus comme vecteurs dans $\mathbb{R}^r$? Ainsi on veut tester

$$(\mathrm{H}) : Q = Q_0 \;\text{ contre }\; (\mathrm{A}) : Q \neq Q_0.$$

On observe un n-échantillon $\mathbf{X}$ de loi Q et on compte le nombre de fois qu'on rencontre la valeur a_j, $j = 1, \ldots, r$:

$$N_j = \sum_{i=1}^n \mathbb{1}_{\{X_i = a_j\}} \;\text{ le nombre des } X_i \text{ égaux à } a_j.$$

On peut voir que $\mathbf{N} = (N_1, \ldots, N_r)^*$ a la loi multimomiale $\mathcal{M}(n, p_1, \ldots, p_r)$:

$$\mathrm{P}(\mathbf{N} = \mathbf{n}) = \mathrm{P}((N_1, \ldots, N_r) = (n_1, \ldots, n_r)) = \frac{n!}{n_1! \ldots n_r!} p_1^{n_1} \ldots p_r^{n_r}, \; n_1 + \ldots + n_r = n.$$

Sa densité par rapport à la mesure de comptage sur $\mathbb{R}^r$ est

$$f_{\mathbf{p}}(\mathbf{n}) = \frac{n!}{n_1! \ldots n_r!} p_1^{n_1} \ldots p_r^{n_r} \mathbb{1}_{\{n_1 + \ldots + n_r = n\}}(\mathbf{n}).$$

Par la méthode des multiplicateurs de Lagrange il est facile de trouver l'estimateur de maximum de vraisemblance de $\mathbf{p}$ avec la contrainte $\hat{p}_1 + \ldots + \hat{p}_r = 1$:

$$\hat{p}_j = \frac{N_j}{n}, \; j = 1, \ldots, r \quad \text{(les fréquences empiriques).}$$

Par la loi forte des grands nombres on sait que lorsque $n \to \infty$,

$$\hat{p}_j \to p_j = \mathrm{E}(\mathbb{1}_{\{X = a_j\}}) = \mathrm{P}(X = a_j), \;\; \text{p.s.}$$

Pour dire si $\mathbf{p} = \pi$ il faut regarder la distance entre ces deux vecteurs. Si on prend n grand il suffira de regarder la distance entre le vecteur des fréquences $\mathbf{f} = (f_1, \ldots, f_r) = (\frac{n_1}{n}, \ldots, \frac{n_r}{n})$ et $(\pi_1, \ldots, \pi_r)$.

Pearson a introduit la “distance” suivante (pas symétrique)

$$\rho(\mathbf{f}, \pi) = \sum_{j=1}^r \frac{(f_j - \pi_j)^2}{\pi_j}.$$

Nous avons à étudier la quantité :

$$T_n = n\rho(\hat{\mathbf{p}}, \pi) = n \sum_{j=1}^r \frac{(\hat{p}_j - \pi_j)^2}{\pi_j} = \sum_{j=1}^r \frac{(N_j - n\pi_j)^2}{n\pi_j}$$

qui doit être proche de zéro sous (H).

Il est clair que sous (A), c'est-à-dire si $\mathbf{p} \neq \pi$, $\rho(\hat{\mathbf{p}}, \pi) \to \rho(\mathbf{p}, \pi)$ p.s. donc dans ce cas $T_n \to \infty$ p.s.

Théorème 2.8 *Soient $\pi_1, \ldots, \pi_r > 0$. Alors, sous* (H) *la suite T_n converge en loi, quand $n \to \infty$ vers la loi $\chi^2(r-1)$. Sous* (A) *la suite T_n tend presque sûrement vers ∞.*

Preuve. On voit que $\mathbf{N} = \sum_{i=1}^n \mathbf{Z}_i$ où

$$\mathbf{Z}_i = \begin{pmatrix} Z_{i,1} \\ \vdots \\ Z_{i,r} \end{pmatrix} = \begin{pmatrix} \mathbb{1}_{\{X_i = a_1\}} \\ \vdots \\ \mathbb{1}_{\{X_i = a_r\}} \end{pmatrix}.$$

sont i.i.d.

Sous (H) l'espérance commune est π est la matrice de covariance est Γ, avec

$$\Gamma_{jk} = \mathrm{Cov}(\mathbf{Z}_1)_{jk} = \mathrm{Cov}(Z_{1,j}, Z_{1,k}) = \delta_{jk}\pi_j - \pi_j\pi_k = \begin{cases} -\pi_j\pi_k, & \text{si } j \neq k \\ \pi_j(1-\pi_j), & \text{si } j = k. \end{cases}$$

Alors, par le théorème central limite, lorsque $n \to \infty$,

$$\frac{1}{\sqrt{n}}(\mathbf{N} - n\pi) \to G \sim \mathcal{N}_r(0, \Gamma) \text{ en loi.}$$

On note D la matrice diagonale $\mathrm{diag}(\frac{1}{\sqrt{\pi_1}}, \ldots, \frac{1}{\sqrt{\pi_r}})$, donc

$$D\frac{1}{\sqrt{n}}(\mathbf{N} - n\pi) \to DG \text{ en loi.}$$

Par calcul

$$D\frac{1}{\sqrt{n}}(\mathbf{N} - n\pi) = \begin{pmatrix} \frac{N_1 - n\pi_1}{\sqrt{n\pi_1}} \\ \vdots \\ \frac{N_r - n\pi_r}{\sqrt{n\pi_r}} \end{pmatrix}$$

et

$$DG \sim \mathcal{N}_r(0, D\Gamma D),$$

où la matrice

$$\Sigma = D\Gamma D = (\delta_{jk} - \sqrt{\pi_j\pi_k})_{jk} = I_r - (\sqrt{\pi})(\sqrt{\pi})^*, \text{ et } (\sqrt{\pi}) = \begin{pmatrix} \sqrt{\pi_1} \\ \vdots \\ \sqrt{\pi_r} \end{pmatrix}.$$

On peut constater que $\|(\sqrt{\pi})\|^2 = 1$, d'où $\Sigma^2 = \Sigma^* = \Sigma$ est une matrice orthogonale et il s'agit de la matrice de la projection orthogonale sur $V := \mathrm{Vect}(\sqrt{\pi})^\perp$. De plus si $\mathbf{Y} \sim \mathcal{N}_r(0, I_r)$, alors $\Sigma\mathbf{Y} \sim \mathcal{N}_r(0, I_r)$.

Ainsi, sous (H), lorsque $n \to \infty$,

$$\begin{pmatrix} \frac{N_1 - n\pi_1}{\sqrt{n\pi_1}} \\ \vdots \\ \frac{N_r - n\pi_r}{\sqrt{n\pi_r}} \end{pmatrix} \to \Sigma \mathbf{Y} \text{ en loi,}$$

d'où, par le théorème de Cochran,

$$n\rho(\hat{\mathbf{p}}, \pi) \to \|\Sigma \mathbf{Y}\|^2 = \|\Pi_V(\mathbf{Y})\|^2 \sim \chi^2(r-1) \text{ en loi,}$$

car $\dim V = r - 1$. □

On peut construire un test **asymptotique**. On choisit c tel que

$$\mathrm{P}(Y > c) = \alpha,\ Y \sim \chi^2(r-1).$$

Le test est, **pour n grand**, donné par

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } t_n > c \\ 0, & \text{sinon,} \end{cases}$$

où $t_n = \sum_{j=1}^r \frac{(n_j - n\pi_j)^2}{n\pi_j}$ est l'observation de T_n. Ce test marche bien si $n \geqslant 30$ et $n\pi_j \geqslant 5$, $j = 1, \dots, r$.

Remarque : Si on veut tester dans le cas où la loi Q_0 est diffuse (par exemple à densité) on choisit d'abord $r \in \mathbb{N}^*$ et $\zeta : \mathbb{R}^d \to \{1, \dots, r\}$, donnée par

$$\zeta(x) = j \text{ si } x \in V_j,\ j = 1, \dots, r,$$

où $\mathbb{R}^d = V_1 \cup \ldots \cup V_r$ est une partition de $\mathbb{R}^d$. On pose alors

$$\pi_j = \mathrm{P}(\zeta(X) = j) = \mathrm{P}(X \in V_j) = Q_0(V_j),\ j = 1, \dots, r.$$

On reprend la même idée que ci-dessus avec, par exemple

$$N_j = \sum_{i=1}^n \mathbb{1}_{\{X_i \in V_j\}}, \text{ etc.}$$

La méthode se généralise pour la situation suivante : soit $\mathbf{X}$ un n-échantillon de loi $Q(\theta) = \sum_{j=1}^r p_j(\theta)\delta_{a_j}$, où les a_j sont connus mais $\theta \in \mathbb{R}^k$, $k \leqslant r$, est un paramètre inconnu. On veut tester

$$\text{(H) : la loi de l'échantillon est dans } \mathcal{P}_0 = \{Q(\theta) : \theta \in \Theta_0\}$$

contre

$$\text{(A) : la loi de l'échantillon n'est pas dans } \mathcal{P}_0.$$

On reprend les fréquences empiriques $\hat{p}_j = \frac{N_j}{n}$, mais la quantité

$$T_n(\theta) = n\rho(\hat{\mathbf{p}}, \mathbf{p}(\theta)) = \sum_{j=1}^{r} \frac{(N_j - np_j(\theta))^2}{np_j(\theta)}$$

n'est plus une statistique (elle dépend de θ). Le rôle de π_j est pris ici de $\{p_j(\theta) : \theta \in \Theta_0\}$. On doit estimer θ par l'estimateur de maximum de vraisemblance. On a le résultat suivant (admis) :

Théorème 2.9 *Supposons que Θ_0 est un ouvert de $\mathbb{R}^k$, $k < r$, que $p_j : \Theta_0 \to [0,1]$ est de classe C^2, $p_j(\theta) \neq 0$, $\forall \theta \in \Theta_0$, que $\forall \theta \in \Theta_0$ la matrice $(\frac{\partial p_j}{\partial \theta_l})_{j,l}$ est de rang maximal k. Supposons que l'estimateur de maximum de vraisemblance $\hat{\theta}_n$ existe pour tout n, et que pour tout n et j, $p_j(\hat{\theta}_n) > 0$. Alors sous* (H), *lorsque $n \to \infty$*

$$T_n(\hat{\theta}_n) \to \chi^2(r-k-1) \ \textit{en loi.}$$

Exemples : On peut construire à l'aide de ce modèle deux test de χ^2 :
1) test d'indépendance : on observe un couple $X = (Y, Z)$ de support $\{1, \ldots, r\} \times \{1, \ldots, s\}$ et on veut tester si Y et Z sont indépendantes, c'est-à-dire si $\mathrm{P}_{(Y,Z)} = \mathrm{P}_Y \otimes \mathrm{P}_Z$. On peut voir qu'ici

$$\mathcal{P}_0 = \{p_{ij}(\theta) = u_i v_j, i = 1, \ldots, r, j = 1, \ldots, s, \theta = (u, v) \in \Theta_0\}$$

où

$$\Theta_0 = \left\{ (u_i, v_j)_{i=1,\ldots,r-1, j=1,\ldots,s-1} \in \mathbb{R}^{r+s-2}, \sum_{i=1}^{r-1} u_i < 1, \sum_{j=1}^{s-1} v_j < 1 \right\}.$$

La dimension de Θ_0 est $r+s-2$ et la loi qui apparaît du théorème est la loi $\chi^2((r-1)(s-1))$. On peut calculer les estimateurs de maximum de vraisemblance :

$$\hat{u}_i = \frac{N_{i\bullet}}{n},\ \hat{v}_j = \frac{N_{\bullet j}}{n},\ \hat{p}_{ij} = \frac{N_{ij}}{n},\ p_{ij}(\hat{\theta}) = \frac{N_{i\bullet} N_{\bullet j}}{n^2},$$

où N_{ij} est le nombre d'observations de (i, j) parmi les n observation.
2) test de symétrie : On observe un couple $X = (Y, Z)$ de support $\{1, \ldots, r\}^2$ et on veut tester si $(Y, Z) \sim (Z, Y)$. A nouveau on peut décrire $\mathcal{P}_0$, Θ_0 (qui a la dimension $\frac{r(r+1)}{2} - 1$) et obtenir la loi $\chi^2(\frac{r(r-1)}{2})$.

2.7.2 Test de Kolmogorov-Smirnov

On se place cette fois-ci sur $\mathbb{R}$ et on regarde la fonction de répartition F plutôt que la loi. On veut tester

$$(\mathrm{H}) : F = F_0 \ \text{ contre } \ (\mathrm{A}) : F \neq F_0$$

où F_0 est une fonction de répartition connue (donnée).

L'idée est d'estimer la fonction F par la fonction de répartition empirique

$$\hat{F}_n(t) = \hat{\mathrm{P}}_n(]-\infty, t]) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{]-\infty,t]}(X_i) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \leqslant t\}}.$$

et de la comparer à F_0. On utilise la distance donnée par la norme uniforme entre F et F_0 ou entre $\hat{F}_n$ et F_0, sous (H).

Proposition 2.4 *Soit F une fonction de répartition et on note* **l'inverse généralisée** *de F*

$$\forall u \in [0,1]\ G(u) = \inf\{t : F(t) \geqslant u\}.$$

1. *Soit $U \sim \mathcal{U}_{[0,1]}$. Alors $G(U)$ a pour fonction de répartition F.*
2. *Soit X une variable aléatoire de fonction de répartition continue F. Alors $F(X)$ est de loi uniforme sur $[0,1]$.*

Preuve. Il est facile de voir que

$$G(u) \leqslant t \leftrightharpoons u \leqslant F(t).$$

1. On en déduit que $\mathrm{P}(G(U) \leqslant t) = \mathrm{P}(U \leqslant F(t)) = F(t)$.
2. Il est clair que $\mathrm{P}(F(X) \leqslant t) = 0$ si $t \leqslant 0$ et $= 1$ si $t > 1$ car F est à valeurs dans [0,1]. Soit $t \in]0,1]$:

$$\mathrm{P}(F(X) < t) = \mathrm{P}(X < G(t)) = \mathrm{P}(X \leqslant G(t)) = F(G(t)) = t, \text{ car } F \text{ est continue.}$$

□

Soient $X_{(1)} < \ldots < X_{(n)}$ les statistiques d'ordre de l'échantillon. Alors on a

$$\hat{F}_n(t) = \begin{cases} 0, & \text{si } t < X_{(1)} \\ \ldots & \ldots \\ \frac{i}{n}, & \text{si } X_{(i)} \leqslant t < X_{(i+1)} \\ \ldots & \ldots \\ 1, & \text{si } t \geqslant X_{(n)} \end{cases}$$

On peut poser $X_{(0)} = -\infty$ et $X_{(n+1)} = \infty$ et alors $\hat{F}_n(t) = \frac{i}{n}$, $\forall t \in [X_{(i)}, X_{(i+1)}[$. On introduit les statistiques suivantes :

$$D_n^+ = \sqrt{n}\sup_{t\in\mathbb{R}}(\hat{F}_n(t) - F_0(t))$$

$$D_n^- = \sqrt{n}\sup_{t\in\mathbb{R}}(F_0(t) - \hat{F}_n(t))$$

et

$$D_n = \sqrt{n} \sup_{t \in \mathbb{R}} |\hat{F}_n(t) - F_0(t)|$$

Il s'agit de mesurer l'adéquation de la fonction de répartition empirique à la fonction F_0 à l'aide la distance de Kolmogorov-Smirnov (norme uniforme).

Proposition 2.5 *Supposons que F_0 est continue et qu'on est sous* (H)*. Alors D_n^+ et D_n^- ont la même loi. De plus les lois de D_n^+ et D_n ne dépendent pas de F_0. Précisement*

$$D_n^+ \sim \sqrt{n} \sup_{0<t<1} (\hat{U}_n(t) - t)$$

et

$$D_n \sim \sqrt{n} \sup_{0<t<1} |\hat{U}_n(t) - t|,$$

où $\hat{U}_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{U_i \leqslant t\}}$ est la fonction de répartition empirique d'un n-échantillon $U_1, \ldots, U_n$ de loi uniforme sur $]0, 1[$.

Preuve. On a, comme F_0 est continue et croissante, sous (H),

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \leqslant t\}} = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{F_0(X_i) \leqslant F_0(t)\}} \sim \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{U_i \leqslant F_0(t)\}} = \hat{U}_n(F_0(t)).$$

Alors

$$D_n^+ = \sqrt{n} \sup_{t \in \mathbb{R}} (\hat{F}_n(t) - F_0(t)) \sim \sqrt{n} \sup_{t \in \mathbb{R}} (\hat{U}_n(F_0(t)) - F_0(t)) = \sqrt{n} \sup_{0<s<1} (\hat{U}_n(s) - s)$$

et

$$D_n = \sqrt{n} \sup_{t \in \mathbb{R}} |\hat{F}_n(t) - F_0(t)| \sim \sqrt{n} \sup_{t \in \mathbb{R}} |\hat{U}_n(F_0(t)) - F_0(t)| = \sqrt{n} \sup_{0<s<1} |\hat{U}_n(s) - s|.$$

Une autre manière, plus intuitive, de voir cette propriété est comme suit : on a, comme F_0 est continue et croissante,

$$D_n^+ = \sqrt{n} \sup_{t \in \mathbb{R}} (\hat{F}_n(t) - F_0(t)) = \sqrt{n} \max_{i=0,\ldots,n} \sup_{t \in [X_{(i)}, X_{(i+1)}[} \left(\frac{i}{n} - F_0(t)\right)$$

$$= \sqrt{n} \max_{i=0,\ldots,n} \left(\frac{i}{n} - F_0(X_{(i)})\right) = \sqrt{n} \max\left\{0, \max_{i=1,\ldots,n} \left(\frac{i}{n} - F_0(X_{(i)})\right)\right\}.$$

Mais on sait que, sous (H), $U_i = F_0(X_i) \sim \mathcal{U}_{[0,1]}$ donc la suite $U_1, \ldots, U_n$ est un n-échantillon de loi uniforme. Comme F_0 est croissante, $F_0(X_1), \ldots, F_0(X_n)$ n'est rien d'autre que le réarrangement par ordre croissant de $U_1, \ldots, U_n$. Ainsi

$$D_n^+ \sim \sqrt{n} \max\left\{0, \max_{i=1,\ldots,n} \left(\frac{i}{n} - U_{(i)}\right)\right\} = \sqrt{n} \max_{i=0,\ldots,n} \left(\frac{i}{n} - U_{(i)}\right)$$

$$= \sqrt{n} \max_{i=0,\ldots,n} \sup_{t\in[U_{(i)},U_{(i+1)}[} \left(\frac{i}{n} - t\right) = \sqrt{n} \sup_{0<t<1} (\hat{U}_n(t) - t).$$

Pour la première partie on voit que

$$D_n^+ \sim \sqrt{n} \sup_{0<t<1} \left(\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{U_i \leqslant t\}} - t \right)$$

$$= \sqrt{n} \sup_{0<t<1} \left(1 - t - \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{1-U_i \leqslant 1-t\}} \right)$$

$$= \sqrt{n} \sup_{0<t'<1} \left(t' - \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{1-U_i \leqslant t'\}} \right) \sim D_n^-$$

car $1 - U \sim U \sim \mathcal{U}_{]0,1[}$. □

Remarque : On rappelle que pour $t \in]0,1[$ fixé, $\mathbb{1}_{]-\infty,t]}(U_i) \sim \mathcal{B}(1,t)$, et que ces variables sont i.i.d. Donc, par le théorème central limite, lorsque $n \to \infty$,

$$\sqrt{n}(\hat{U}_n(t) - t) \to \mathcal{N}(0, t(1-t)) \text{ en loi.}$$

Ensuite si on fixe $t_1, \ldots, t_d \in]0,1[$, par la version vectorielle du théorème central limite, quand $n \to \infty$,

$$\sqrt{n} \begin{pmatrix} \hat{U}_n(t_1) - t_1 \\ \vdots \\ \hat{U}_n(t_d) - t_d \end{pmatrix} \to \mathcal{N}_d(0, \Gamma) \text{ en loi}$$

où,

$$\Gamma_{jk} = \mathrm{E}\left\{ (\mathbb{1}_{]-\infty,t_j]}(U) - t_j)(\mathbb{1}_{]-\infty,t_k]}(U) - t_k) \right\} = t_j \wedge t_k - t_j t_k,\ j,k = 1, \ldots, d.$$

On peut montrer que la famille de variables aléatoires (processus)

$$\left(\sqrt{n}(\hat{U}_n(t) - t) : t \in]0,1[\right)$$

converge en loi vers une famille de variables aléatoires gaussiennes (processus gaussien) ayant les covariances données par la fonction $(s,t) \mapsto s \wedge t - st$.

On en déduit le résultat suivant (admis) :

Théorème 2.10 *(de Kolmogorov-Smirnov)*
Sous (H)*, lorsque* $n \to \infty$,

$$\mathrm{P}(D_n^+ > x) \to e^{-2x^2} \quad \textit{et} \quad \mathrm{P}(D_n > x) \to 2 \sum_{k \geqslant 1} (-1)^{k+1} e^{-2k^2x^2}.$$

En fait, pour n petit les valeurs de $\sqrt{n}(\hat{U}_n(t) - t)$ sont tabulées et sinon, on utilise le théorème précedent pour trouver les fonctions de répartition.

Si on veut tester

$$(\mathrm{H}) : F = F_0 \text{ contre } (\mathrm{A}) : F \neq F_0$$

avec F_0 continue, on choisit c tel que

$$2 \sum_{k \geqslant 1} (-1)^{k+1} e^{-2k^2c^2} = \alpha$$

et le test est **asymptotique** : **pour** n **grand**,

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } d_n > c \\ 0, & \text{sinon,} \end{cases}$$

où d_n est l'observation de D_n. Si on veut tester

$$(\mathrm{H}) : F = F_0 \text{ contre } (\mathrm{A}) : F > F_0$$

avec F_0 continue, on choisit $c = \sqrt{\ln \frac{1}{\sqrt{\alpha}}}$ et le test est, **pour** n **grand**

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } d_n^+ > c \\ 0, & \text{sinon,} \end{cases}$$

où d_n^+ est l'observation de D_n^+. Par exemple, si on observe $x_1, \ldots, x_n$, on les ordonne par ordre croissant $x_{(1)} < \ldots < x_{(n)}$, on calcule

$$d_n^+ = \sqrt{n} \max \left\{ 0, \max_{i=1,\ldots,n} \left(\frac{i}{n} - F_0(x_{(i)}) \right) \right\}$$

et on compare sa valeur à celle de $\sqrt{\ln \frac{1}{\sqrt{\alpha}}}$. Sinon, on calcule

$$\begin{aligned} d_n &= \sqrt{n} \max_{i=1,\ldots,n} \max \left(\frac{i}{n} - F_0(x_{(i)}); F_0(x_{(i+1)}) - \frac{i}{n} \right) \\ &= \sqrt{n} \max \left\{ 0, \max_{i=1,\ldots,n} \max \left(\frac{i}{n} - F_0(x_{(i)}); F_0(x_{(i)}) - \frac{i-1}{n} \right) \right\} \end{aligned}$$

et on compare sa valeur à celle donée par les tables.

2.8 Tests non-paramétriques de comparaison

Soient $(X_1, \ldots, X_n)$ et $(Y_1, \ldots, Y_m)$ deux échantillons de lois Q_X et Q_Y, respectivement. Les deux lois sont inconnues, mais cette fois-ci on veut seulement savoir si $Q_X = Q_Y$ ou non. On donnera ici plusieurs méthodes permettant de répondre à cette question, suivant que les deux échantillons sont indépendants ou pas. Pour simplifier, on supposera que les lois sont sur $\mathbb{R}$ et on utilisera les fonctions de répartition F et G, respectivement, des deux lois.

2.8.1 Comparaison de deux échantillons indépendants

On suppose de plus ici que les deux *échantillons* $\mathbf{X}$ *et* $\mathbf{Y}$ *sont indépendants* et que les *fonctions de répartition* F *et* G *sont continues*. On veut tester

$$(\mathrm{H}) : F = G \text{ contre } (\mathrm{A}) : F \neq G(\text{ ou } F > G \text{ ou } F < G)$$

a. Test de Kolmogorov-Smirnov

On estime F et G par les fonctions de répartition empirique

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leqslant t\}} \text{ et } \hat{G}_m(t) = \frac{1}{m}\sum_{j=1}^{m} \mathbb{1}_{\{Y_j \leqslant t\}}.$$

On va approcher la distance uniforme entre les deux fonctions F et G par la quantité

$$\sup_{t\in\mathbb{R}}(\hat{F}_n(t) - \hat{G}_m(t)).$$

On va montrer que sous (H) la loi de cette variable aléatoire ne dépend pas de $F = G$.

Proposition 2.6 *Soient deux échantillons* $\mathbf{X}$ *et* $\mathbf{Y}$ *indépendants de fonctions de répartition* F *et* G *continues. Si* $F = G$*, les lois des statistiques*

$$D^+_{n,m} = \sqrt{\frac{nm}{n+m}}\sup_{t\in\mathbb{R}}(\hat{F}_n(t) - \hat{G}_m(t))$$

et

$$D_{n,m} = \sqrt{\frac{nm}{n+m}}\sup_{t\in\mathbb{R}}|\hat{F}_n(t) - \hat{G}_m(t)|$$

ne dépendent pas de F.

Preuve. On note $X_{(1)} < \ldots < X_{(n)}$ et $Y_{(1)} < \ldots < Y_{(m)}$ les statistiques d'ordre des échantillons. On a

$$\sup_{t\in\mathbb{R}}(\hat{F}_n(t) - \hat{G}_m(t)) = \max_{i=0,\ldots,n, j=0,\ldots,m}\ \sup_{t\in V_{ij}}\left(\frac{i}{n} - \frac{j}{m}\right),$$

où $V_{ij} := [X_{(i)}, X_{(i+1)}[\cap[Y_{(j)}, Y_{(j+1)}[$ est non-vide si et seulement si $Y_{(j)} \in [X_{(i)}, X_{(i+1)}[$ ou $Y_{(j+1)} \in [X_{(i)}, X_{(i+1)}[$. On en déduit

$$\begin{aligned}\sup_{t\in\mathbb{R}}(\hat{F}_n(t) - \hat{G}_m(t)) &= \sup\left(\frac{i}{n} - \frac{j-1}{m} : (i,j) \text{ t.q. } Y_{(j)} \in [X_{(i)}, X_{(i+1)}[\right)\\ &= \sup\left(\frac{i}{n} - \frac{j-1}{m} : (i,j) \text{ t.q. } F(Y_{(j)}) \in [F(X_{(i)}), F(X_{(i+1)})[\right)\end{aligned}$$

puisque F est croissante et continue. On note $U_i = F(X_i)$ et $V_j = G(Y_j)$. Sous (H), $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ est un échantillon de taille $n+m$ de fonction de répartition F continue. Alors $(U_1, \ldots, U_n, V_1, \ldots, V_m)$ est $n+m$-échantillon de loi uniforme sur $[0,1]$. Comme F est croissante :

$$\sup_{t\in\mathbb{R}}(\hat{F}_n(t)-\hat{G}_m(t)) \sim \sup\left(\frac{i}{n} - \frac{j-1}{m} : (i,j) \text{ t.q. } V_{(j)} \in [U_{(i)}, U_{(i+1)}[\right) = \sup_{t\in\mathbb{R}}(\hat{U}_n(t)-\hat{V}_m(t))$$

où $\hat{U}_n$ et $\hat{V}_m$ sont les fonctions de répartition empirique des échantillons indépendants $\mathbf{U}$ et $\mathbf{V}$ de loi uniforme sur $[0,1]$. □

On admet le résultat suivant :

Théorème 2.11 *Sous* (H)*, lorsque* $n, m \to \infty$ *de telle sorte que* $a \leqslant \frac{n}{m} \leqslant b$,

$$\mathrm{P}(D_{n,m}+ > x) \to e^{-2x^2} \quad \textit{et} \quad \mathrm{P}(D_{n,m} > x) \to 2\sum_{k\geqslant 1}(-1)^{k+1}e^{-2k^2x^2}.$$

Si on veut tester

$$\text{(H)} : F = G \ \text{ contre } \ \text{(A)} : F \neq G$$

lorsque F et G sont continues, on choisit c tel que

$$2\sum_{k\geqslant 1}(-1)^{k+1}e^{-2k^2c^2} = \alpha$$

et le test est **asymptotique** : **pour** n **et** m **grands**,

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } d_{n,m} > c \\ 0, & \text{sinon,} \end{cases}$$

où $d_{n,m}$ est l'observation de $D_{n,m}$.

b. Test de Wilcoxon-Mann-Whitney de la somme des rangs

On note cette fois-ci $(Z_1, \ldots, Z_{n+m}) = (X_1, \ldots, X_n, Y_1, \ldots, Y_m)$. Sous (H) $\mathbf{Z}$ est un $n+m$-échantillon de fonction de répartition F continue.

Soit la permutation aléatoire R d'ordre $n+m$ de $(1, \ldots, n+m)$ telle que R_i donne le rang de Z_i dans le réarrangement croissant $Z_{(1)} < \ldots < Z_{(n+m)}$. On note $W = \sum_{i=1}^{n+m} R_i$ la somme des rangs des Z_i dans le réarrangement croissant $Z_{(1)} < \ldots < Z_{(n+m)}$. Le résultat suivant est classique (voir aussi l'exercice 1.11.3-1) :

Proposition 2.7 *La loi de la statistique R est uniforme sur l'ensemble des permutations de $(1, \ldots, n+m)$. En particulier, sa loi ne dépend pas de F.*

Ainsi la loi de W ne dépend pas de F. Pour des petites valeurs de n et m la loi de W est tabulée. De plus :

Théorème 2.12 *Sous* (H),

$$\mathrm{E}(W) = \frac{n(n+m+1)}{2}, \ \mathrm{Var}(W) = \frac{nm(n+m+1)}{12}$$

et, lorsque $n, m \to \infty$ *de telle sorte que* $a \leqslant \frac{n}{m} \leqslant b$,

$$\frac{W - \mathrm{E}(W)}{\sqrt{\mathrm{Var}(W)}} \to \mathcal{N}(0,1) \ \textit{ en loi.}$$

Preuve. On va calculer seulement l'espérance et la variance de W. On a $R_i = 1 + \sum_{j=1}^{n+m} \mathbb{1}_{Z_j < Z_i}$. Alors

$$W = n + \sum_{i=1}^{n}\sum_{j=1}^{n+m} \mathbb{1}_{Z_j<Z_i} = n + \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{1}_{X_j<X_i} + \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbb{1}_{Y_j<X_i} = n + \frac{n(n-1)}{2} + V,$$

où on a noté $V = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbb{1}_{Y_j<X_i}$. Sous (H), Y_j et X_i sont indépendantes et de même loi, donc $\mathrm{P}(Y_j < X_i) = \mathrm{P}(X_i < Y_j) = 1/2$. On en déduit que $\mathrm{E}(V) = \frac{nm}{2}$.

Par ailleurs, $\mathrm{Var}(W) = \mathrm{Var}(V) = \mathrm{E}(V^2) - \mathrm{E}(V)^2$. On a

$$\mathrm{E}(V^2) = \sum_{i,j} \mathrm{P}(Y_j < X_i) + \sum_{i,j,k,\ell,(i,j)\neq(k,\ell)} \mathrm{P}(Y_j < X_i, Y_\ell < X_k).$$

Sous (H),

$$\mathrm{P}(Y_j < X_i, Y_\ell < X_k) = \begin{cases} 1/4 & \text{si } j \neq \ell \text{ et } i \neq k \\ 1/3 & \text{si } j = \ell \text{ ou } i = k \text{ et } (i,j) \neq (k,\ell). \end{cases}$$

On obtient

$$\mathrm{E}(V^2) = \frac{nm}{2} + \frac{nm(n-1) + nm(m-1)}{3} + \frac{nm(n-1)(m-1)}{4} = \frac{nm(3nm+n+m+1)}{12}.$$

□

Si on veut tester

$$(\mathrm{H}) : F = G \ \text{ contre } \ (\mathrm{A}) : F \neq G$$

lorsque F et G sont continues, on choisit

$$c = \sqrt{\frac{nm(n+m+1)}{12}} g_{\alpha/2}$$

et le test est **asymptotique** : **pour** n **et** m **grands**,

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } |w - n(n+m+1)/2| > c \\ 0, & \text{sinon,} \end{cases}$$

où w est l'observation de W.

2.8.2 Comparaison de deux échantillons appariés

Deux échantillons $(X_1, \dots, X_n)$ et $(Y_1, \dots, Y_n)$ sont dits **appariés** s'ils ont la même taille et les couples aléatoires (X_i, Y_i) sont indépendants. Il s'agit le plus souvent de l'observation du même caractère à deux moments différents. Il n'y a pas indépendance entre X_i et Y_i, en général.

On peut, par exemple, penser que X_i est le poids de la i-ème personne avant un régime et Y_i le poids de la même personne après le régime. Il n'y a pas d'indépendance entre X_i et Y_i mais il y a indépendance entre les poids entre deux personnes. Il est naturel de se demander si le régime est efficace ou pas (c'est-à-dire si la proportion de personnes ayant maigri est de plus de 50%.

On veut tester

$$(\mathrm{H}) : \mathrm{P}(Y < X) \geqslant \frac{1}{2} \text{ contre } (\mathrm{A}) : \mathrm{P}(Y < X) < \frac{1}{2}.$$

Soit $Z_i = X_i - Y_i$ et soit m une médiane de Z_i définie par $\mathrm{P}(Z_i \leqslant m) = \frac{1}{2}$. On continue de supposer que les fonctions de répartition des X_i et de Y_i sont continues (et celle de Z_i aussi), donc la médiane existe toujours mais n'est pas forcément unique. L'hypothèse (H) est alors : Z_i admet une médiane $m \geqslant 0$. Ainsi on fait des tests pour la médiane d'une loi.

a. Test du signe

Soit $\mathbf{Z}$ un n-échantillon de loi de fonction de répartition F inconnue mais supposée continue. On veut tester

$$(\mathrm{H}) : \exists m \text{ médiane de } F \text{ t.q. } m \geqslant m_0 \text{ contre } (\mathrm{A}) : \forall m \text{ médiane de } F,\ m < m_0,$$

où m_0 est une valeur fixée (donnée). On considère la statistique

$$S_n = \sum_{i=1}^{n} \mathbb{1}_{Z_i \leqslant m_0}.$$

Alors $S_n \sim \mathcal{B}(n, p)$, où $p = \mathrm{P}(Z_1 \leqslant m_0)$. Sous (H) $p \geqslant \mathrm{P}(Z_1 \leqslant m) = \frac{1}{2}$. Le test est alors donné par

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } s_n \geqslant k \\ 0, & \text{sinon,} \end{cases}$$

où s_n est l'observation de S_n et où k est tel que

$$\alpha \geqslant \sup_{p \leqslant 1/2} \mathrm{P}_p(X \geqslant k),\ X \sim \mathcal{B}(n, p).$$

Pour k fixé la fonction $p \mapsto \mathrm{P}_p(X \geqslant k)$ est croissante donc k est donné par

$$\alpha \geqslant \sup_{p \leqslant 1/2} \mathrm{P}_p(S_n \geqslant k) = \mathrm{P}_{1/2}(S_n \geqslant k).$$

b. Test de signe et rang

Revenons l'exemple précédent. On peut apprécier l'efficacité du régime autrement : il se peut qu'il y ait moins de 50/des gens qui maigrissent en suivant le régime, mais ceux qui maigrissent, maigrissent beaucoup. Dans l'hypothèse où le régime n'a pas d'effets, la loi de $Z_i = X_i - Y_i$ devrait être symétrique par rapport à 0. Dans le cas où le régime a un effet amaigrissant, la loi de Z_i devrait accorder un poids plus important aux valeurs positives.

Soit $\mathbf{Z}$ un n-échantillon de loi de fonction de répartition F continue. On veut tester

$$(\mathrm{H}): \ F \text{ est symétrique contre } (\mathrm{A}): \ F \text{ charge les valeurs positives.}$$

Ainsi, sous (H) le régime n'est pas efficace et sous (A) le régime est efficace.

On range les $|Z_i|$ par ordre croissant et on note par R_i le rang de $|Z_i|$ dans le réarrangement. Considérons la statistique

$$T^+ = \sum_{i:Z_i \geqslant 0} R(i)$$

la somme des rangs des Z_i positifs. Sous (A) T^+ a tendance à être plus grand que sous (H). On a besoin de la loi de T^+ sous (H).

Proposition 2.8 *Sous* (H), $T^+ \sim \sum_{j=1}^n j\xi_j$, *où* $(\xi_1, \ldots, \xi_n)$ *est un n-échantillon de loi* $\mathcal{B}(1, \frac{1}{2})$.

Preuve. Notons ε_i le signe de Z_i. Ainsi $\varepsilon_i = 1$ si $Z_i \geqslant 0$ et $\varepsilon_i = -1$ si $Z_i < 0$. La loi de Z_i est symétrique par rapport à 0, alors ε_i et $|Z_i|$ sont deux variables aléatoires indépendantes. En effet, par symétrie :

$$\begin{aligned}\mathrm{P}(\varepsilon_i = 1, |Z_i| \geqslant t) &= \mathrm{P}(Z_i \geqslant 0, |Z_i| \geqslant t) = \mathrm{P}(Z_i \geqslant t)\\ &= \mathrm{P}(-Z_i \geqslant t) = \mathrm{P}(Z_i \geqslant -t) = \mathrm{P}(\varepsilon_i = -1, |Z_i| \geqslant t).\end{aligned}$$

On en déduit que

$$\mathrm{P}(\varepsilon_i = 1, |Z_i| \geqslant t) = \frac{1}{2}\mathrm{P}(|Z_i| \geqslant t) = \mathrm{P}(\varepsilon_i = 1)\mathrm{P}(|Z_i| \geqslant t),$$

$$\mathrm{P}(\varepsilon_i = -1, |Z_i| \geqslant t) = \frac{1}{2}\mathrm{P}(|Z_i| \geqslant t) = \mathrm{P}(\varepsilon_i = -1)\mathrm{P}(|Z_i| \geqslant t).$$

Soit D la permutation inverse de R. D est définie par $|Z_{D(1)}| < \ldots < |Z_{D(n)}|$ et ne dépend donc que de $|Z_1|, \ldots, |Z_n|$. Par conséquent, D est indépendante de $(\varepsilon_1, \ldots, \varepsilon_n)$. On en déduit que $(\varepsilon_{D(1)}, \ldots, \varepsilon_{D(n)})$ a même loi que $(\varepsilon_1, \ldots, \varepsilon_n)$.

Par ailleurs,

$$T^+ = \sum_{i=1}^n R(i)\mathbb{1}_{Z_i \geqslant 0} = \sum_{j=1}^n j\mathbb{1}_{Z_{D(j)} \geqslant 0} = \sum_{j=1}^n j\mathbb{1}_{\varepsilon_{D(j)}=1} \sim \sum_{j=1}^n j\mathbb{1}_{\varepsilon_j=1}.$$

En posant $\xi_j = \mathbb{1}_{\varepsilon_j=1}$, $(\xi_1, \dots, \xi_n)$ est un n-échantillon de loi de Bernoulli de paramétre $1/2$. □

La loi de T^+ est tabulée pour n petit. Pour n grand ($n \geqslant 20$), on pourra utiliser le résultat suivant :

Théorème 2.13 *Sous* (H),

$$\mathrm{E}(T^+) = \frac{n(n+1)}{4}, \ \mathrm{Var}(T^+) = \frac{n(n+1)(2n+1)}{24}$$

et, lorsque $n \to \infty$,

$$\frac{T+ - \mathrm{E}(T^+)}{\sqrt{\mathrm{Var}(T^+)}} \to \mathcal{N}(0,1) \ \text{ en loi.}$$

Le test est donné par

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{si } t^+ > c + \frac{n(n+1)}{4} \\ 0, & \text{sinon,} \end{cases}$$

où t^+ est l'observation de T^+ et où c est $c = \sqrt{\frac{n(n+1)(2n+1)}{24}} g_{\alpha/2}$.

2.9 Exercices

2.9.1 Tests statistiques : construire et comparer

1. Soit $(X_1, \dots, X_n)$ n-échantillon de loi $\mathcal{N}(\theta, 1)$ et soient $\theta_0 > \theta_1$. Construire le test de Neyman-Pearson pour tester H : $\theta = \theta_0$ contre A : $\theta = \theta_1$. Indiquer le test de niveau $\alpha \in]0,1[$ pour $\theta = \theta_0$. Application : $\theta_0 = 10$, $n = 25$ et $\alpha = 0,05$. Que vaut la puissance de ce test pour $\theta_1 = 9$?

2. Soit $(X_1, \dots, X_n)$ n-échantillon de loi $\mathcal{N}(\theta, 1)$ et soit θ_0 un réel fixé connu. On veut tester H : $\theta \leqslant \theta_0$ contre A : $\theta > \theta_0$.

a) Montrer que le test de rapport de vraisemblance est

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{si } \sqrt{n}(\overline{x} - \theta_0) > c \\ 0 & \text{sinon,} \end{cases}$$

où c est une constante.

b) écrire la fonction puissance $\beta : \mathbb{R} \to [0,1]$ de ce test et faire son graphe. Montrer que le test est sans biais.

c) On veut que $\beta(\theta_0) = 0,1$ et que $\beta(\theta_0 + 1) = 0,8$. Montrer qu'alors $c = 1,28$ et $n = 5$.

d) On ne fixe plus n. Trouver c pour que le test soit de niveau $\alpha \in]0,1[$ pour $\theta \leqslant \theta_0$. S'agit-il d'un test u.p.p. contre $\theta > \theta_0$?

e) Application : $\theta_0 = 9, n = 25$ et $\alpha = 0,05$. Que vaut la puissance de ce test en $\theta_1 = 10$?

3. Soit $(X_1, \ldots, X_n)$ n-échantillon de loi $\mathcal{N}(\theta, 1)$ et soit θ_0 un réel fixé connu. On veut tester H : $\theta \geqslant \theta_0$ contre A : $\theta < \theta_0$. Montrer que le test de rapport de vraisemblance est

$$\psi(\mathbf{x}) = \begin{cases} 1 & \text{si } \overline{x} < \theta_0 - g_\alpha/\sqrt{n} \\ 0 & \text{sinon,} \end{cases}$$

où g_α est telle que $\mathrm{P}(G > g_\alpha) = \alpha$, avec G variable gaussienne standard. écrire la fonction puissance de ce test et montrer que ce test est de niveau α pour $\theta \geqslant \theta_0$ et u.p.p. contre $\theta < \theta_0$.

4. Soit $(X_1, \ldots, X_n)$ n-échantillon de loi $\mathcal{N}(\theta, 1)$ et soit θ_0 un réel fixé connu. On veut construire des tests pour tester H : $\theta = \theta_0$ contre A : $\theta \neq \theta_0$.

a) On écrit $\{\theta = \theta_0\} = \{\theta \leqslant \theta_0\} \cap \{\theta \geqslant \theta_0\}$. Utiliser l'exercice **2** pour construire un test de $\theta \leqslant \theta_0$ contre $\theta > \theta_0$ avec une constante c' et ensuite l'exercice **3** pour construire un test de $\theta \geqslant \theta_0$ contre $\theta < \theta_0$ avec une constante c''. En déduire que la région de rejet du test recherché est $\{\sqrt{n}(\overline{x} - \theta_0) > c'\} \cup \{\sqrt{n}(\overline{x} - \theta_0) < c''\}$. Que devient ce test lorsqu'on pose $c' = -c'' > 0$?

b) Construire directement le test du rapport de vraisemblance ϕ et comparer avec le test trouvé au point précédent. Quel est le choix de la constante pour que le test soit de niveau α pour $\theta = \theta_0$.

c) Montrer par calcul direct que ce test satisfait les conditions suivantes $\mathrm{E}_{\theta_0}[\phi(\mathbf{X})] = \alpha$ et $\mathrm{E}_{\theta_0}[\overline{X}\phi(\mathbf{X})] = \alpha \mathrm{E}_{\theta_0}[\overline{X}]$. En déduire par le résultat du cours qu'il s'agit d'un test sans biais de niveau α pour $\theta = \theta_0$ u.p.p. contre $\theta \neq \theta_0$. Dessiner le graphe de sa fonction puissance.

5. Soit $(X_1, \ldots, X_n)$ n-échantillon de loi $\mathcal{N}(\theta, \sigma^2)$. On suppose que σ^2 est inconnu. Trouver les tests de rapport de vraisemblance pour tester respectivement

a) H : $\theta \leqslant \theta_0$ contre A : $\theta > \theta_0$;

b) H : $\theta \geqslant \theta_0$ contre A : $\theta < \theta_0$;

c) H : $\theta = \theta_0$ contre A : $\theta \neq \theta_0$.

Préciser ces tests pour qu'ils soient de niveau $\alpha = 0,05$ lorsque $n = 25$.

6. Soit $(X_1, \ldots, X_n)$ n-échantillon de loi $\mathcal{N}(1, \theta)$. Construire les tests de rapport de vraisemblance pour tester respectivement

a) H : $\theta \leqslant \theta_0$ contre A : $\theta > \theta_0$;

b) H : $\theta \geqslant \theta_0$ contre A : $\theta < \theta_0$;

c) H : $\theta = \theta_0$ contre A : $\theta \neq \theta_0$.

Préciser ces tests pour qu'ils soient de niveau $\alpha = 0,05$ lorsque $n = 25$.

7. Soit $(X_1, \dots, X_r)$ r-échantillon de loi $\mathcal{N}(m_1, \sigma_1^2)$ et $(Y_1, \dots, Y_s)$ s-échantillon de loi $\mathcal{N}(m_2, \sigma_2^2)$. Les deux échantillons sont indépendants.

a) On pose $\theta = (m_1, m_2, \sigma_1^2, \sigma_2^2) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+^* \times \mathbb{R}_+^*$. Construire le test de rapport de vraisemblance pour tester H :$\theta \in \Theta_0$ contre $\theta \in \Theta_0^c$, où $\Theta_0 = \{\theta \in \Theta : \sigma_1^2 = \sigma_2^2\}$.

b) On suppose cette fois-ci qu'en effet, $\sigma_1^2 = \sigma_2^2 =: \sigma^2$ mais inconnu. Le paramètre sera dans ce cas $\theta = (m_1, m_2, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+^*$. Construire le test de rapport de vraisemblance pour tester H :$\theta \in \Theta_0$ contre $\theta \in \Theta_0^c$, où $\Theta_0 = \{\theta \in \Theta : m_1 = m_2\}$.

Préciser ces tests pour qu'ils soient de niveau $\alpha = 0,1$ lorsque $n = 10$.

8. Soit $(X_1, \dots, X_n)$ un n-échantillon de loi normale $\mathcal{N}(0, 1/\theta)$.

a) Trouver $\hat{\theta}^*$ l'estimateur de maximum de vraisemblance de θ. Construire un estimateur sans biais de θ. Cet estimateur est-il efficace, asymptotiquement efficace ?

b) Construire un test de $\theta = 1$ contre $\theta > 1$. Si $n = 15$ et si $\sum_{i=1}^{15} x_i^2 = 6,8$, effectuer le test niveau 5% pour $\theta = 1$. Calculer approximativement la puissance de ce test en $\theta = 2, 3, 4, 5$ à l'aide des tables de lois et dessiner la fonction puissance. Ce test est-il u.p.p. contre $\theta > 1$?

c) Construire le test u.p.p. contre $\theta \neq 1$ au niveau 5% pour $\theta = 1$ pour les mêmes données.

9. Soit $(X_1, \dots, X_n)$ un n-échantillon de loi de densité

$$f(x, a) = (1/a)\exp(-x/a)\mathbb{1}_{[0,\infty[}(x),\ a > 0.$$

a) Déterminer l'estimateur de maximum de vraisemblance de a. Cet estimateur est-il efficace ?

b) Quel est le test u.p.p. contre $a > a_0$ au niveau α pour $a \leqslant a_0$

c) Calculer approximativement la région de rejet de ce test pour $a_0 = 1$, $n = 50$ et $\alpha = 0,05$. Quelle est la puissance au point $a = 2$?

d) Dans les mêmes conditions, si on désire une puissance de 0,95 pour la contre $a = 2$, quelle taille d'échantillon faut-il prévoir ?

10. Soit $p \in \mathbb{R}_+^*$. On observe un n-échantillon de loi de densité

$$f(x, p) = p^2 x \exp(-px)\mathbb{1}_{\{x>0\}}.$$

a) Quel est le test le plus puissant contre $p = p_1$ de niveau α pour $p = p_0$?

b) Quel test proposez-vous pour tester $p = p_0$ contre $p > p_0$ ou $p \neq p_0$?

11. On nourrit 20 petits cochons : la moitié d'entre eux reçoivent leur ration de protéines sous forme de cacahuètes crues et l'autre sous forme de cacahuètes grillées. On veut savoir si le fait de griller les cacahuètes a un effet sur l'augmentation de poids des petits cochons. L'augmentation du poids (en centaines de grammes) des animaux mesurée au bout d'une semaine est (première ligne pour le cas "crues", x, la deuxième pour le cas "grillées", y) :

x	62	56	61	58	60	44	56	60	56	63
y	53	51	62	55	59	56	61	54	47	57

.

a) Tester l'égalité des variances dans les deux échantillons au niveau 5%. Préciser les hypothèses.

b) En "acceptant" l'égalité des variances, peut-on considérer que le fait de griller les cacahuètes a une influence sur la croissance des petits cochons (on fera un test au niveau 5%).

2.9.2 Tests statistiques : modèle linéaire, tests non-paramétriques

1. On observe les variables aléatoires $X_i = \beta + \gamma t_i + \varepsilon_i$, $i = 1, \ldots, n$, où $(\varepsilon_1, \ldots, \varepsilon_n)$ est un n-échantillon de loi $\mathcal{N}(0, \sigma^2)$ et $t_1, \ldots, t_n$ sont des réels connus. On veut estimer le paramètre $\theta = (\beta, \gamma, \sigma^2)$.

a) Construire l'estimateur du maximum de vraisemblance pour $\hat{\theta}^* = (\widehat{\beta^*}, \widehat{\gamma^*}, \widehat{\sigma^{2,*}})$ de $\theta = (\beta, \gamma, \sigma^2)$.

b) Montrer que $\widehat{\beta^*}$ et $\widehat{\gamma^*}$ sont sans biais et calculer leur variances. Trouver leurs lois. Montrer que si $\overline{t} = \frac{1}{n}\sum_{i=1}^{n} = 0$ alors $\widehat{\beta^*}$ et $\widehat{\gamma^*}$ sont non-corrélées.

c) Construire le test de rapport de vraisemblance pour $\beta = 0$ contre $\beta \neq 0$ de niveau $\alpha \in]0, 1[$. Construire le test de rapport de vraisemblance pour $\gamma = 0$ contre $\gamma \neq 0$ de niveau $\alpha \in]0, 1[$. Application : $n = 27$ et $\alpha = 0, 05$.

d) écrire des intervalles de confiance pour β, pour γ de coefficients de sécurité $1 - \alpha \in]0, 1[$. Construire une région de confiance pour (β, γ) de coefficient de sécurité $1 - \alpha \in]0, 1[$. Application : $n = 27$ et $\alpha = 0, 05$.

2. Appliquer les résultats de l'exercice précédent pour les données suivantes :

t	5	10	15	20	25	30
x	0,10	0,21	0,30	0,35	0,44	0,62

et $\alpha = 0, 05$. Pouvez-vous prédire la valeur de X en $t_0 = 17$.

3. On observe les variables aléatoires $X_i = \beta + \gamma(t_i - \overline{t}) + \varepsilon_i$, $i = 1, \ldots, n$, où $(\varepsilon_1, \ldots, \varepsilon_n)$ est un n-échantillon de loi $\mathcal{N}(0, \sigma^2)$, $t_1, \ldots, t_n$ sont des réels connus et $\overline{t} = \frac{1}{n}\sum_{i=1}^{n} = 0$.

a) Construire le test de rapport de vraisemblance pour $\gamma = 0$ contre $\gamma \neq 0$ de niveau $\alpha \in]0,1[$.

b) écrire des intervalles de confiance pour γ de coefficients de sécurité $1-\alpha \in]0,1[$.

4. Appliquer les résultats de l'exercice précédent pour les données suivantes :

t	3,80	3,72	3,67	3,60	3,54
x	1,36	1,23	1,09	0,82	0,61

et $\alpha = 0,05$. Pouvez-vous prédire la valeur de X en $t_0 = 17$.

5. Un dé est lancé 600 fois. On observe que les faces apparaissent un certain nombre de fois indiqué dans le tableau suivant :

face	1	2	3	4	5	6
nbr. fois	100	94	103	89	110	104

Au niveau $\alpha = 0,1$ tester si le dé est truqué.

6. Une espèce de tulipes peut avoir trois couleurs : rouge, rose et violet. L'expérience montre que ces couleurs apparaissent avec probabilités, respectivement $\frac{1}{12}$, $\frac{3}{12}$ et $\frac{8}{12}$. Un statisticien commande 60 tulipes au hasard et il reçoit 6 rouges, 18 roses et 36 violets. Tester la validité du modèle au niveau $\alpha = 0,05$.

7. Le devoir de statistique d'un groupe sont notés de 0 à 100 mais on rend seulement des qualificatifs : TB pour points entre 90 et 100, B entre 75 et 89, M entre 60 et 74, F entre 50 et 59 et enfin I pour points entre 0 et 49. Le correcteur aimerait avoir une répartition du nombre de points obtenus de loi gaussienne de paramètres 75 et 81. Après avoir corrigé il trouve :

qualificatifs	TB	B	M	F	I
nbr. fois	3	12	10	4	1

Tester l'hypothèse du correcteur au niveau $\alpha = 0,05$.

8. Même question pour le tableau suivant :

$n \leqslant 90$	$90 < n \leqslant 100$	$100 < n \leqslant 110$	$110 < n \leqslant 120$	$120 < n \leqslant 130$	$n > 130$
10	18	23	22	18	9

Pour estimer m et σ^2 on pourra prendre les points au milieu des intervalles ainsi que 65 pour le premier intervalle et 160 pour le dernier.

9. On observe les temps de panne d'une machine. On utilise un appareil qui donne en fait les logarithmes de ces temps de panne. Les observations sont : 2,88 ; 3,36 ; 3,50 ; 3,73 ; 3,74 ; 3,82 ; 3,88 ; 3,95 ; 3,95 ; 3,99 ; 4,02 ; 4,22 ; 4,23 ; 4,23 ; 4,23 ; 4,43 ; 4,53 ; 4,59 ; 4,66 ; 4,66 ; 4,85 ; 4,85, 5,16.

a) On veut tester à l'aide d'un test de χ^2 l'hypothèse que la loi des logarithmes des temps de panne est une loi normale de paramètres $\log(50) = 3,912$ et variance 0,25. On pourra décomposer $\mathbb{R}$ en intervalles ayant chacun la probabilité 1/4. Montrer que pour la loi $\mathcal{N}(3,912; 0,25)$ il s'agit des intervalles $]-\infty; 3,575]$, $]3,575; 3,912]$, $]3,912; 4,249]$ et $]4,249; \infty[$. Appliquer le test de χ^2 à un niveau plus petit que 0,3, ainsi qu'à un niveau plus grand que 0,4.

b) On reprend les mêmes données mais cette fois-ci on va tester que la loi est une loi normale, mais sans connaître les paramètres. Montrer que 4,150 et 0,2722 sont des estimations sans biais pour l'espérance et la variance. On décompose $\mathbb{R}$ avec les mêmes intervalles qu'au point précédent.Calculer les probabilités de ces intervalles pour la nouvelle loi et appliquer ensuite le test de χ^2 à un niveau plus petit que 0,27. Préciser le nombre de degrés de liberté.

c) Utiliser le test de Kolmogorov-Smirnov pour tester l'hypothèse du point **a)**, c'est-à-dire que les observations proviennent d'une loi $\mathcal{N}(3,912; 0,25)$.

10. Un magasin vend deux types de céréales pour le petit déjeuner A et B. Le propriétaire observe que les hommes et les femmes n'achètent pas au hasard et il veut savoir s'il y a indépendance entre le sexe de l'acheteur et la marque de céréales achetée. Il bâtit le tableau suivant :

	A	B
Homme	9	6
Femme	13	16

Construire un test de χ^2 d'indépendance et conclure.

11. Deux traitements sont appliqués et on note trois types de résultats :

	guérison	amélioration	état stationnaire
traitement A	280	210	110
traitement B	220	90	90

Construire un test de χ^2 d'homogénéité pour étudier si les traitements sont différents ?

12. On observe un caractère 21 fois. On veut tester si ce caractère est de loi normale de paramètres 2 et 1 par le test de Kolmogorov-Smirnov : 0,3 ; 0,7 ; 0,9 ; 1,2 ; 1,4 ; 1,4 ; 1,5 ; 1,5 ; 1,6 ; 1,9 ; 2,0 ; 2,1 ; 2,1 ; 2,3 ; 2,5 ; 2,6 ; 2,7 ; 3,0 ; 3,8 ; 3,9 ; 4,0.

Chapitre 3

Sujets d'examens 2002-2005

Sujet d'examen partiel
Maîtrise de Mathématiques - Statistiques
mercredi 3 avril 2002 - durée 3 heures - calculatrice interdite
documents (cours manuscrit ou polycopié) autorisés

Exercice I.

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de même loi de densité $f_\theta(x) = c_\theta\,(\theta - x)\mathbb{1}_{]0,\theta[}(x)$, $\theta \in \Theta =]0, \infty[$. Que valent c_θ, $\mathrm{E}(X_1)$, $\mathrm{Var}(X_1)$ et $\mathrm{Var}(X_1^2)$?

1) Construire deux estimateurs $\theta^*_{0,n}$ et $\theta^*_{1,n}$ du paramètre θ par la méthode des moments.

2) $\theta^*_{0,n}$ et $\theta^*_{1,n}$ sont-ils consistants ? asymptotiquement normaux ?

3) Lequel de ces deux estimateurs est asymptotiquement meilleur ?

Exercice II.

La durée de vie d'un mécanisme est une variable aléatoire positive T. Pour chaque instant $t > 0$, son taux de panne est $r(t) := \lim_{h\downarrow 0} \mathrm{P}(T \leqslant t + h \mid T > t)/h$.

1) Dans la fiche technique on trouve, pour $t > 0$, $r(t) = 1/(2\theta\sqrt{t})$, où $\theta > 0$. Calculer la densité f_θ de support $]0, \infty[$ (c'est-à-dire nulle sur $]-\infty, 0]$) de la loi de T (on pourra exprimer le taux de panne r à l'aide de la fonction de répartition de T).

2) Montrer que la statistique $(\sqrt{T_1} + \ldots + \sqrt{T_n})/n$ associée à un n-échantillon de loi de densité f_θ, est un estimateur de $\theta > 0$ par la méthode de maximum de vraisemblance, qu'il est consistant, asymptotiquement normal, exhaustif, complet, sans biais, efficace et R-efficace.

Exercice III.

Soit X le chiffre obtenu lors du lancer d'un dé (éventuellement truqué) : la probabilité d'obtenir la face j est $\theta_j \in]0, 1[$ (inconnu), $j = 1, \ldots, 6$ et $\theta_1 + \ldots + \theta_6 = 1$. On veut estimer le paramètre vectoriel $\theta = (\theta_1, \ldots, \theta_6)^\star \in \Theta = \{\theta \in \mathbb{R}^6 : \theta_j > 0,\ \theta_1 + \ldots + \theta_6 = 1\}$.

On effectue n lancers, on note $X_1, \ldots, X_n$ les chiffres obtenus au cours de ces lancers et pour $j = 1, \ldots, 6$ on note N_j le nombre des X_i égaux à j. Le vecteur aléatoire $(N_1, \ldots, N_6)^\star$ sera noté par N.

On veut étudier la statistique N, obtenir un estimateur de θ et étudier ses propriétés.

1) a) Décrire la loi des X_i.

b) Montrer que $N_j = \sum_{i=1}^n Z_{i,j}$, où $Z_{i,j} = \mathbb{1}_{\{X_i = j\}}$, $j = 1, \ldots, 6$. En déduire que N est de loi multinomiale $\mathcal{M}(n, \theta_1, \ldots, \theta_6)$.

c) Soit μ la mesure de comptage sur $\mathbb{R}^6$ (donc $\mu(B)$ donne le nombre de points de B à coordonnées entières). Montrer que la densité de N par rapport à μ est :

$$f_\theta(n_1, \ldots, n_6) = \frac{n!}{n_1! \ldots n_6!} \theta_1^{n_1} \ldots \theta_6^{n_6} \mathbb{1}_A(n),$$

$$A = \{(n_1, \ldots, n_6)^\star \in \mathbb{R}^6 : n_j \geqslant 0, \text{ entiers}, j = 1, \ldots, 6, n_1 + \ldots + n_6 = n\}.$$

d) Montrer que la loi de N est de type exponentiel. En déduire que N est une statistique suffisante complète pour θ. Qu'en est-il de la statistique $(N_1, \ldots, N_5)^\star$?

2) a) Pour $i = 1, \ldots, n$ on note Z_i le vecteur aléatoire $(Z_{i,1}, \ldots, Z_{i,6})^\star$. Montrer que $Z_1, \ldots, Z_n$ sont des vecteurs aléatoires indépendants et de même loi. Calculer le vecteur espérance $\mathrm{E}(Z_1)$ et la matrice de covariance $\mathrm{Cov}(Z_1)$. Décrire la loi des Z_i à l'aide de la base canonique $\{e_1, \ldots, e_6\}$ sur $\mathbb{R}^6$.

b) Vérifier que N/n est la moyenne empirique associée à l'échantillon $Z_1, \ldots, Z_n$. En déduire que N/n est un estimateur sans biais du paramètre θ.

c) En utilisant la méthode des multiplicateurs de Lagrange, montrer que le seul point critique de la fonction vraisemblance soumis à la contrainte $\theta_1 + \ldots + \theta_6 = 1$, est $\theta_j^* = n_j/n$, $j = 1, \ldots, 6$. En déduire que N/n est l'estimateur du maximum de vraisemblance du paramètre θ.

d) Montrer que N/n est un estimateur consistant et asymptotiquement normal de θ. Quel est son paramètre de dispersion ?

Exercice IV.

Soit $\theta \in \Theta \subset \mathbb{R}$ le paramètre d'une famille dominée de lois. On construit, à l'aide d'un n-échantillon, θ_n^* un estimateur de θ asymptotiquement normal, de paramètre de dispersion $\sigma^2(\theta)$. Notons Y une variable aléatoire de loi de $\chi^2(1)$.

1) a) Exprimer la limite en loi de $n(\theta_n^* - \theta)^2$, quand $n \to \infty$, à l'aide de Y.

b) Montrer que θ_n^* est un estimateur consistant.

2) Soit $g : \mathbb{R} \to \mathbb{R}$ une fonction de classe C^2 avec g'' bornée.

a) Montrer que, si $g'(\theta) \neq 0$, $g(\theta_n^*)$ est un estimateur de $g(\theta)$ asymptotiquement normal, de paramètre de dispersion $\sigma^2(\theta)[g'(\theta)]^2$.

(On admettra le résultat suivant pour des variables aléatoires réelles : si ξ_n converge en loi vers ξ et si η_n converge en probabilité vers la constante réelle a, alors $\xi_n \eta_n$ converge en loi vers $a\xi$).

b) Supposons maintenant que $g'(\theta) = 0$, mais $g''(\theta) \neq 0$. Montrer que, lorsque $n \to \infty$,

$$\sqrt{n}\,[g(\theta_n^*) - g(\theta)] \xrightarrow{\text{prob}} 0 \ \text{ et } \ n\,[g(\theta_n^*) - g(\theta)] \xrightarrow{\text{loi}} \frac{1}{2}\sigma^2(\theta) g''(\theta)\, Y.$$

Sujet d'examen

Maîtrise de Mathématiques - Statistiques
lundi 10 juin 2002 - durée 3 heures
documents (cours manuscrit ou polycopié) et calculatrice autorisés

Exercice I.

On note par f_0 la densité de probabilité de la loi uniforme $\mathcal{U}_{[0,1]}$ et par f_1 la densité de probabilité triangulaire sur $[0,1]$ (c'est-à-dire $f_1(x) = 4x$ pour $0 \leqslant x < \frac{1}{2}$, $f_1(x) = 4 - 4x$ pour $\frac{1}{2} \leqslant x \leqslant 1$ et 0 ailleurs). On observe une seule fois une variable aléatoire X de densité f. Construire le test le plus puissant de l'hypothèse $f = f_0$ contre $f = f_1$ au niveau $\alpha = 0,05$.

Exercice II.

Une usine fabrique des pellicules photographiques. On mesure l'épaisseur des couches sensibles de ces pellicules sur un échantillon de 616 pellicules. On répartit les épaisseurs par tranche de 1 micron et on désigne par x_i l'épaisseur moyenne de la classe i. Les épaisseurs inférieures à 2,5 microns sont regroupées dans une seule classe et de même pour les épaisseurs supérieurs à 7,5 microns Les autres observations forment cinq classes :

épaisseur des couches sensibles (classe no. i)	nombre de pellicules (n_i)	épaisseur moyenne (x_i)	probabilités gaussiennes (π_i)
moins de 2,5	8	2	
de 2,5 à 3,5	47	3	
de 3,5 à 4,5	121	4	
de 4,5 à 5,5	219	5	
de 5,5 à 6,5	149	6	
de 6,5 à 7,5	60	7	
plus de 7,5	12	8	

Les données suggèrent que l'épaisseur des couches sensibles pourrait suivre une loi gaussienne.

1. Rappeler les estimateurs du maximum de vraisemblance de l'espérance m et de la variance σ^2 pour la loi gaussienne $\mathcal{N}(m, \sigma^2)$. Sont-ils biaisés ?
2. Calculer les observations $\bar{x}$ et s^2 de la moyenne empirique et de la variance empirique. En déduire des estimations (sans biais) de l'espérance et de la variance d'une gaussienne.
3. Calculer les probabilités π_i pour qu'une variable gaussienne de paramètres égaux aux estimations précédentes se trouve dans la classe i. On utilisera les tables de lois.
4. Quel est le nombre de degrés de liberté de la loi à utiliser ? Pourquoi ?
5. Tester au niveau $\alpha = 0,01$ l'hypothèse que l'épaisseur des couches sensibles peut être ajustée à une loi gaussienne dont on indiquera les paramètres.

Exercice III.

Soit $X_1, \dots, X_n$ variables aléatoires indépendantes de même loi uniforme $\mathcal{U}_{[0,\theta]}$, $\theta > 0$.

1. Trouver l'estimateur de maximum de vraisemblance $\hat{\theta}^*$. S'agit-il d'une statistique suffisante ? Montrer que la densité de probabilité de la fonction pivotale $T_n(\theta) = \hat{\theta}^*/\theta$ est $h_n(x) = nx^{n-1}\mathbb{1}_{[0,1]}(x)$.
2. Soit ξ une variable aléatoire de densité de probabilité h_n et on définit deux réels a, b, $0 \leqslant a \leqslant b \leqslant 1$ tels que $\mathrm{P}(a \leqslant \xi \leqslant b) = 1 - \alpha$. Quelle relation satisfont a, b et α ?
3. Construire à l'aide de a, b et $\hat{\theta}^*$ un intervalle de confiance pour θ de coefficient de sécurité $1 - \alpha$. Quelle est sa longueur L ?
4. On regarde a et L comme deux fonctions de b, $a = a(b)$ et $L = L(b)$. Utiliser l'égalité qui relie a et b pour calculer da/db. En déduire une expression de la dérivée dL/db. Étudier la monotonie de la fonction L.
5. Pour quelle valeur de b la fonction L atteint son minimum ? En déduire la valeur de a et enfin l'intervalle de confiance de longueur minimale. Quelle la longueur moyenne de l'intervalle obtenu ?

Exercice IV.

On met en service en parallèle n composants électroniques identiques. Leurs durées de vie $T_1, \dots, T_n$ sont supposées indépendantes de même loi exponentielle de paramètre $1/2\theta$, $\theta > 0$. On note $T_{(1)} < \dots < T_{(n)}$ les statistiques d'ordre associées à l'échantillon $(T_1, \dots, T_n)$. Des impératifs économiques incitent à observer **uniquement** les instants des r premières pannes, c'est-à-dire $(T_{(1)}, \dots, T_{(r)})$ (on posera $T_{(0)} = 0$). On définit $Y_i = T_{(i)} - T_{(i-1)}$, pour $1 \leqslant i \leqslant n$. Il est clair qu'il suffit d'observer $(Y_1, \dots, Y_r)$.

1. Montrer que $Y_i \sim \mathcal{E}(\frac{n-i+1}{2\theta})$, $1 \leqslant i \leqslant n$ et que les variables $Y_1, \dots, Y_n$ sont indépendantes. En déduire que la densité de $(Y_1, \dots, Y_r)$ par rapport à la mesure $\mu(dy_1, \dots, dy_r) = \mathbb{1}_{\{y_1>0,\dots,y_r>0\}} dy_1 \dots dy_r$ sur $\mathbb{R}^r$, est de la forme $c_{r,n}(2\theta)^{-r} \exp[-s_r/(2\theta)]$, où $c_{r,n}$ est une constante positive et $s_r = \sum_{i=1}^r (n-i+1)y_i$.
2. Trouver l'estimateur du maximum de vraisemblance $\hat{\theta}^*$ de θ et l'exprimer en termes de $S_r = \sum_{i=1}^r (n-i+1)Y_i$. Montrer que S_r/θ suit une loi $\chi^2(2r)$ et que $\hat{\theta}^*$ est sans biais.
3. On suppose que r et n tendent vers $+\infty$. Montrer que l'estimateur du maximum de vraisemblance $\hat{\theta}^*$ est consistant (on pourra montrer que l'estimateur est consistant en moyenne quadratique). Que peut-on dire de la limite en loi de $\sqrt{r}(\hat{\theta}^* - \theta)$?
4. On suppose que $r = 4$ et soit $\theta_0 = 1000$. Justifier pourquoi le test donné par $\phi(y_1, \dots, y_r) = 1$, si $s_r < c$ et $= 0$ sinon, est un test u.p.p. de niveau $\alpha = 5\%$ pour tester $\theta \geqslant \theta_0$ contre $\theta < \theta_0$. Déterminer c et une valeur approchée de sa puissance en $\theta_1 = 500$?
5. On garde les mêmes valeurs pour θ_0 et θ_1 et le même niveau. Déterminer r tel que la puissance en θ_1 soit de 95% (on pourra utiliser l'approximation gaussienne $\mathcal{N}(k, 2k)$

de la loi $\chi^2(k)$ quand k est assez grand pour trouver une approximation de la quantile $c_{k,\alpha}$ d'ordre α de $\chi^2(k)$ à l'aide de la quantile g_α d'ordre α de la loi gaussienne standard).

Sujet d'examen de septembre

Maîtrise de Mathématiques - Statistiques
mercredi 4 septembre 2002 - durée 3 heures
documents (cours manuscrit ou polycopié) et calculatrice autorisés

Exercice I.
Soit X_1, X_2, X_3 un 3-échantillon de loi de fonction caractéristique $(1 - it\theta)^{-1}$.

1) Calculer à partir de la fonction caractéristique l'espérance et la variance de cette loi.

2) Pourquoi X_2 est un estimateur sans biais de θ ? Calculer sa variance.

3) Montrer que $S = X_1 + X_2 + X_3$ est une statistique exhaustive pour θ. Est-elle complète ? minimale ?

4) Montrer que $\mathrm{E}_\theta(X_1 \mid S) = \mathrm{E}_\theta(X_2 \mid S) = \mathrm{E}_\theta(X_3 \mid S)$. En déduire la valeur commune de ces espérances conditionnelles.

5) Indiquer un estimateur efficace θ^* pour θ. Justifier pourquoi X_2 n'est pas un estimateur efficace. Comparer les estimateurs θ^* et X_2. θ^* est-il R-efficace ?

6) Trouver la loi de la fonction pivotale $T_3(\theta) = \frac{2}{\theta}S$. En déduire qu'il existe deux réels $0 < a < b$ tels que $\mathrm{P}(a \leqslant \eta \leqslant b) = 0,95$, où η est une variable aléatoire ayant la même loi que $T_3(\theta)$. Si $a = 1,237$ quelle sera la valeur pour b ?

7) En déduire un intervalle de confiance pour θ de coefficient de sécurité 0,95.

8) Trouver le test u.p.p. pour tester $\theta \geqslant \theta_0 = 1000$ contre $\theta < 1000$ au niveau 5%.

9) Quelle taille doit avoir l'échantillon telle que la puissance au point $\theta_1 = 105,5$ soit 95% ?

Exercice II.
Le tableau ci-dessous donne la répartition du nombre de jours (f_i) sans accident, avec un accident, ..., avec quatre accidents pour une période de 50 jours, dans une ville :

nombre d'accidents (n_i)	nombre de jours (f_i)
0	21 jours
1	18 jours
2	7 jours
3	3 jours
4	1 jour

1. Calculer le nombre moyen $\bar{x}$ d'accidents par jour et la variance empirique s^2 du nombre d'accidents par jour à une décimale (ne pas faire d'approximations). Observer l'égalité de ces deux valeurs approximatives.

2. Le résultat précédent nous fait penser que la loi (inconnue) du nombre d'accidents pourrait être une loi de Poisson. Soit η variable aléatoire de Poisson. Calculer les probabilités $\mathrm{P}(\eta = i)$, $i = 0, \ldots, 4$, lorsque le paramètre prend la valeur trouvée au point précédent.

3. On veut vérifier l'ajustement des observations à la loi de Poisson par un test de χ^2. Expliquer le nombre de degrés de liberté utilisé pour la loi de χ^2. Donner la forme du test et la décision.

Exercice III.
Sur une cible de surface 1 ayant la forme d'un disque, la partie centrale (noire) est de rayon $\sqrt{p/\pi}$, $0 < p < 1$. On fait une suite de tirs indépendants sur la cible jusqu'à ce qu'on touche exactement r fois la partie centrale (r est un entier non nul donné) et on désigne par $T^{(r)}$ la variable aléatoire égale au rang du tir qui touche la cible pour la r-ième fois.

1) **a)** Expliquer pourquoi les variables aléatoires $T^{(1)}, T^{(2)} - T^{(1)}, \ldots, T^{(r)} - T^{(r-1)}$ sont indépendantes et de même loi de fonction génératrice donnée par $G(s) = \frac{ps}{1-(1-p)s}$, $|s| \leqslant 1$. En déduire la fonction génératrice de $T^{(r)}$, son espérance m et sa variance σ^2. (La fonction génératrice d'une variable aléatoire ζ est $G_\zeta(s) = \mathrm{E}(s^\zeta)$.)

b) Montrer que pour $k \geqslant r$, $\mathrm{P}(T^{(r)} = k) = \mathrm{C}_{k-1}^{r-1} p^r (1-p)^{k-r}$.

2) **a)** Que vaut la fonction caractéristique $\varphi_{T^{(r)}}$?

b) Soit la variable aléatoire $Z^{(r)} = (T^{(r)} - m)/\sigma$. Montrer que la fonction caractéristique $\varphi_{Z^{(r)}}(t)$ tend vers $e^{-t^2/2}$ lorsque $r \to \infty$. Que signifie-t-il ce résultat ?

c) On note $N_1 := T^{(1)}$ et $N_j = T^{(j)} - T^{(j-1)}$, $j = 2, \ldots, r$. Retrouver sans calcul le résultat précedent en utilisant les variables aléatoires $N_1, \ldots, N_r$.

3) On reprend $N_1, \ldots, N_r$ le r-échantillon trouvé à la question **1a)**.

a) S'agit-il d'un modèle exponentiel ? Indiquer une statistique exhaustive complète. Trouver l'estimateur de maximum de vraisemblance de p.

b) Montrer que pour $r \geqslant 2$, $(r-1)/(T^{(r)} - 1)$ est un estimateur sans biais. On pourra utiliser la valeur de la somme $\sum_{k \geqslant r} \mathrm{C}_{k-1}^{r-1} p^r (1-p)^{k-r}$.

c) Donner un estimateur sans biais de $1/p$ et calculer sa variance. Est-il efficace ? Est-il R-efficace ?

Sujet d'examen partiel

Maîtrise de Mathématiques - Statistiques
samedi 5 avril 2003 - durée 3 heures - calculatrice interdite
documents (cours manuscrit ou polycopié) autorisés

Exercice I. Les cotations d'une valeur en bourse pendant n jours sont des réalisantions de variables aléatoires indépendantes $X_1, \dots, X_n$ de même loi, d'espérance m et de variance σ^2, les deux inconnues.

1) Indiquer un estimateur sans biais θ_0^* de la variance. Est-il consistant ?

2) On veut estimer l'écart-type $\sigma > 0$ de la loi commune des X_i.

a) Pourquoi $g(\theta_0^*) := \sqrt{\theta_0^*}$ est-il un estimateur biaisé de l'écart-type ?

b) Supposons que la loi des cotations est normale $\mathcal{N}(m, \sigma^2)$. Que vaut $\mathrm{E}_{m,\sigma^2}[g(\theta_0^*)]$? Comment faut-il modifier la fonction borélienne g pour que $\theta_0^{**} = g(\theta_0^*)$ soit un estimateur sans biais de l'écart-type ?

3) On continue à supposer que la loi des cotations est $\mathcal{N}(m, \sigma^2)$ et on veut estimer m^2. Considérons les estimateurs suivants : $T_{1,n} = \bar{X}^2$ et $T_{2,n} = \bar{X}^2 - \frac{\theta_0^*}{n}$.

a) Ces estimateurs sont-ils biaisés ? consistants ?

b) Supposons que $m \neq 0$. Étudier la normalité asymptotique de ces estimateurs. Lequel est asymptotiquement meilleur ?

c) Supposons que $m = 0$. Trouver les limites en loi de $\sqrt{n}\, T_{i,n}$ et de $nT_{i,n}$, $i = 1, 2$. Montrer qu'en fait $\sqrt{n}\, T_{i,n}$, $i = 1, 2$, convergent en probabilité.

Exercice II. Une calculatrice possède une fonction "random" qui fournit de nombres "au hasard". Le constructeur garantit que les nombres sont répartis uniformément dans un intervalle $[\alpha, \beta]$ ($\alpha < \beta$) mais on ne précise pas les valeurs de α et β. Il faut donc les estimer à l'aide des observations de n variables aléatoires indépendantes $X_1, \dots, X_n$ de même densité uniforme sur $[\alpha, \beta]$.

1) Vérifier que $(X_{(1)}, X_{(n)})$ est une statistique exhaustive pour (α, β). Montrer (ou admettre) que cette statistique est complète.

2) Montrer que $\frac{1}{n-1}(nX_{(1)} - X_{(n)})$ et $\frac{1}{n-1}(nX_{(n)} - X_{(1)})$ sont des estimateurs sans biais respectivement pour α et pour β. (On pourra calculer d'abord $E_{\alpha,\beta}[X_{(1)}]$ et $E_{\alpha,\beta}[X_{(n)}]$.) En déduire que ce sont les estimateurs efficaces pour α et, respectivement, pour β.

Sur le site internet du constructeur est précisée la valeur $\alpha = 0$. Seul $\beta > 0$ est inconnu.

3) Indiquer l'estimateur de maximum de vraisemblance $\hat{\beta}_n^*$ pour β.

4) Calculer le biais et la qualité quadratique de $\hat{\beta}_n^*$. Que valent les limites, lorsque $n \to \infty$, de $\mathrm{E}_\beta\left[\hat{\beta}_n^*\right]$ et de $\mathrm{E}_\beta\left[n(\hat{\beta}_n^* - \beta)^2\right]$?

5) Quelle est la limite en loi de la suite de variables aléatoires $\{n(\beta - \hat{\beta}_n^*) : n \geqslant 1\}$? Commenter votre résultat.

Exercice III. Lors d'un référendum les journalistes d'un quotidien veulent se faire une idée sur le pourcentage de personnes qui, dans la population française, répondent "oui" à la question posée. Ils interrogent "au hasard" et "avec remise" n personnes. Les réponses sont notées $X_1, \ldots, X_n$.

1) Un des journalistes propose d'utiliser le modèle basé sur la famille paramétrique $\mathcal{P} = \{\mathcal{B}(1,p) : 0 < p < 1\}$. Montrer que cette famille est dominée par la mesure $\mu = \delta_0 + \delta_1$, donner sa densité f_p et montrer qu'il s'agit d'une famille exponentielle.

2) Calculer l'information de Fisher associée à ce modèle.

3) Le journaliste utilise la méthode du maximum de vraisemblance pour trouver un estimateur $\hat{p}_n^*$ du pourcentage. Donner l'expression de $\hat{p}_n^*$. Est-il consistant, asymptotiquement normal (si oui, indiquer son paramètre de dispersion), exhaustif, minimal, complet, efficace, R-efficace ?

4) Un autre journaliste, ayant fait des études de statistique, se propose d'estimer la variance, $g(p) = p(1-p)$, vue comme nouveau paramètre.

a) On note $q = g(p) \Leftrightarrow p = g^{-1}(q)$ et $\check{f}_q(x) = f_{g^{-1}(q)}(x)$. Montrer que, pour le paramètre q, l'information de Fisher $\check{I}(q)$ peut s'exprimer à l'aide de $I(p)$ et de $g'(p)$. On pourra justifier et utiliser l'égalité suivante : $\frac{d}{dq} \ln \check{f}_q(x) = \frac{dp}{dq} \cdot \frac{d}{dp} \ln f_p(x)$ (et aussi que $(g^{-1})' = \frac{1}{g' \circ g^{-1}}$). Quelle est la borne de Cramer-Rao, notée $\frac{r(p)}{n}$, pour les estimateurs sans biais de $g(p)$?

b) Montrer que $T = \bar{X}(1-\bar{X})$ est l'estimateur de la variance obtenu par la méthode des moments. Calculer $\mathrm{E}_p(T)$. Dire pourquoi $\bar{X}$ est une statistique exhaustive pour $g(p)$. En déduire un estimateur sans biais $\widetilde{T}$ de $g(p)$ et montrer qu'il s'agit d'un estimateur efficace.

c) Montrer que : $\mathrm{Var}_p(\widetilde{T}) = \left(\frac{n}{n-1}\right)^2 \mathrm{E}_p\left[(g(\bar{X}) - g(p))^2\right] - \frac{p^2(1-p)^2}{(n-1)^2}$.

Vérifier que : $g(\bar{X}) = g(p) + (\bar{X} - p)(1-2p) - (\bar{X} - p)^2$. En déduire :

$$\mathrm{E}_p\left[(g(\bar{X}) - g(p))^2\right] = (1-2p)^2 \frac{p(1-p)}{n} + \mathrm{E}_p\left[(\bar{X} - p)^4\right] - 2(1-2p)\mathrm{E}_p\left[(\bar{X} - p)^3\right].$$

On note, pour $i = 1, \ldots, n$, $\xi_i = X_i - p$ et $\bar{\xi} = \bar{X} - p$. Vérifier que :

$$\mathrm{E}_p[(\sum_{i=1}^n \xi)^3] = n\mathrm{E}_p\left[\xi_i^3\right] \text{ et } \mathrm{E}_p[(\sum_{i=1}^n \xi)^4] = n\mathrm{E}_p\left[\xi_i^4\right] + C_4^2 n(n-1)\mathrm{E}_p\left[\xi_i^2\right]^2$$

et trouver les valeurs de $\mathrm{E}_p\left[\bar{\xi}^3\right]$, $\mathrm{E}_p\left[\bar{\xi}^4\right]$.

En déduire : $\mathrm{E}_p\left[n(g(\bar{X}) - g(p))^2\right] = r(p)\left(1 - \frac{1}{n}\right)^2 + \frac{6n-5}{n^2} p^2(1-p)^2$.

d) Montrer que, pour tout $p \in]0,1[$, $n\mathrm{Var}_p(\widetilde{T}) > r(p)$, mais $\lim_{n\to\infty} n\mathrm{Var}_p(\widetilde{T}) = r(p)$. Que peut-on conclure ?

Sujet d'examen

Maîtrise de Mathématiques - Statistiques
Mercredi 4 juin 2003 - durée 3 heures
documents (cours manuscrit ou polycopié) et calculatrice autorisés

Exercice I. La durée de vie d'un système est une variable aléatoire X de densité $f_\lambda(x) = e^{\lambda - x}\mathbb{1}_{x \geqslant \lambda}$ (le paramètre $\lambda \geqslant 0$ étant la durée de vie minimale du système). On observe un n-échantillon $(X_1, \ldots, X_n)$, $n \geqslant 2$, de densité f_λ.

1. Trouver l'estimateur $\hat{\lambda}$ de λ par la méthode du maximum de vraisemblance et montrer que $\hat{\lambda}$ est une statistique exhaustive pour λ.
2. Calculer la loi de $\hat{\lambda}$ et étudier s'il s'agit d'une statistique complète. Peut-on construire un estimateur $\lambda^\star$ sans biais en ajoutant une constante à $\hat{\lambda}$? Calculer la variance de $\lambda^\star$.
3. Que vaut $\mathrm{E}(X)$? Utiliser la méthode des moments pour déduire un autre estimateur sans biais $\lambda^\sharp$ de λ. Calculer sa variance.
4. Comparer les estimateurs $\lambda^\star$ et $\lambda^\sharp$. Ces deux estimateurs sont-ils convergents ?
5. à l'aide de λ^* et d'une statistique complète, construire un estimateur efficace pour λ. Est-il unique ? Peut-on étudier son R-efficacité en utilisant l'inégalité de Cramer-Rao ?
6. Quelle est la loi de $\hat{\lambda} - \lambda$? En déduire que $2n(\hat{\lambda} - \lambda) \sim \chi^2(2)$. Indiquer deux valeurs $c_1 < c_2$ telles que $[\hat{\lambda} - \frac{c_2}{n}, \hat{\lambda} - \frac{c_1}{n}]$ soit un intervalle de confiance pour λ de coefficient de sécurité 95%.
7. Si $0 \leqslant a < b$ sont tels que $\mathrm{P}(a \leqslant \hat{\lambda} - \lambda \leqslant b) = 95\%$, construire un autre intervalle de confiance pour λ de coefficient de sécurité 95%. En déduire l'intervalle de confiance de longueur minimale. On pourra étudier la fonction $b = b(a)$ comme fonction de a.
8. Quelle est la limite en loi de $\sqrt{n}(\lambda^\sharp - \lambda)$, lorsque $n \to \infty$? Est-il possible d'en déduire un intervalle de confiance asymptotique pour λ de coefficient de sécurité 95% ?

Exercice II. On veut tester la validité d'un compteur Geiger-Muller et dans ce but on mesure le nombre x de désintégrations qui se sont produites durant des intervalles temps de longueur 30 secondes. Les 100 déterminations successives ont donné : une fois 0 désintégrations, 4 fois une désintégration, 12 fois deux désintégrations, ... On résume tout cela dans le tableau suivant :

nombre de désintégrations	0	1	2	3	4	5	6	7	8	9	10	$\geqslant 11$
nombre de fois	1	4	12	18	22	17	11	6	4	3	2	0

On désigne par X la variable aléatoire comptant le nombre de désintégrations. Les données suggèrent que X pourrait suivre une loi de Poisson. Pour confirmer ou infirmer cette hypothèse, on va réaliser un test de chi-deux. Doit-on regrouper des valeurs ? Donner un estimateur sans biais du paramètre de la loi et calculer sa réalisation. Comment calculer les probabilités théoriques ? Combien de degrés de liberté doit-on utiliser ? Construire le test et conclure.

Exercice III. On observe deux échantillons indépendants $\mathbf{X} = (X_1, \ldots, X_n)$ de loi $\mathcal{E}(\frac{1}{\theta})$ et $\mathbf{Y} = (Y_1, \ldots, Y_m)$ de loi $\mathcal{E}(\frac{1}{\theta'})$. Les paramètres $\theta > 0$ et $\theta' > 0$ sont inconnus. On veut comparer les paramètres, plus précisement on veut tester (H) : $\theta = \theta'$ contre (A) : $\theta \neq \theta'$.

1. Préciser la vraisemblance du $(n+m)$-échantillon avec le paramètre $(\theta, \theta') \in \Theta$, ainsi que l'ensemble Θ_0 asocié à l'hypothèse (H).
2. On note $R(\mathbf{x}, \mathbf{y}) := \dfrac{\sup_{(\theta,\theta')\in\Theta_0} f_{\theta,\theta'}(\mathbf{x}, \mathbf{y})}{\sup_{(\theta,\theta')\in\Theta} f_{\theta,\theta'}(\mathbf{x}, \mathbf{y})}$. Trouver pour quelles valeurs de θ et θ' ces supremums sont atteints. On pose $\bar{x} := \frac{1}{n}\sum_{i=1}^n x_i$, $\bar{y} := \frac{1}{m}\sum_{j=1}^m y_j$ et $t(\mathbf{x}, \mathbf{y}) := \frac{n\bar{x}}{n\bar{x}+m\bar{y}}$. Montrer que $R(\mathbf{x}, \mathbf{y}) = \dfrac{(n+m)^{n+m}}{n^n m^m} t(\mathbf{x}, \mathbf{y})^n (1 - t(\mathbf{x}, \mathbf{y}))^m$.
3. Peut-on construire un test pour (H) à l'aide de $R(\mathbf{x}, \mathbf{y})$? étudier la variation de la fonction $t \mapsto t^n(1-t)^m$ sur $[0, 1]$, et montrer que le test est de la forme
$$\phi(\mathbf{x}, \mathbf{y}) = \begin{cases} 1, & \text{si } t(\mathbf{x}, \mathbf{y}) < c_1 \text{ ou } t(\mathbf{x}, \mathbf{y}) > c_2 \\ 0, & \text{sinon.} \end{cases}$$
Préciser les conditions satisfaites par les constantes c_1 et c_2. On pourra faire un dessin.
4. Trouver les lois de $\sum_{i=1}^n X_i$, $\sum_{j=1}^n Y_j$ et ensuite de $T = t(\mathbf{X}, \mathbf{Y})$ sous (H). Expliquer le choix des constantes c_1 et c_2 pour construire un test pour (H) de niveau $\alpha \in]0, 1[$.
5. On admet que l'hypothèse (H) est vérifiée, c'est-à-dire que $\theta = \theta'$ et on veut estimer θ. Utiliser d'abord l'échantillon $\mathbf{X}$ et ensuite l'échantillon $\mathbf{Y}$ pour construire (par la méthode de votre choix) deux estimateurs sans biais pour θ. Parmi les combinaisons linéaires de ces deux estimateurs trouver l'estimateur sans biais et de variance minimale.

Sujet d'examen de septembre

Maîtrise de Mathématiques - Statistiques
jeudi 4 septembre 2003 - durée 3 heures
documents (cours manuscrit ou polycopié) et calculatrice autorisés

Exercice I. On observe le nombre de clients devant un guichet, à la même heure, durant n jours consécutifs : $x_1, \dots, x_n$. Ces valeurs sont les réalisations des variables aléatoires $X_1, \dots, X_n$, indépendantes, ayant la même loi de Poisson de paramètre $\lambda > 0$ inconnu.

1. Indiquer deux estimateurs de λ par la méthode des moments. Sont-ils convergents ?

2. Montrer que parmi les deux estimateurs précédents celui qui est sans biais est aussi l'estimateur du maximum de vraisemblance $\hat{\lambda}$ pour l'échantillon. S'agit-il d'un estimateur efficace ? R-efficace ? exhaustif ? complet ?

3. Que vaut la limite en loi de $\sqrt{n}(\hat{\lambda} - \lambda)$? Trouver une fonction borélienne $g : [0, \infty[\to [0, \infty[$ telle que la limite de la suite de variables aléatoires $\sqrt{n}(g(\hat{\lambda}) - g(\lambda))$ soit indépendante de λ. On pourra d'abord supposer que g satisfait des bonnes hypothèses de régularité, trouver g et ensuite vérifier la régularité supposée.

4. Utiliser le résultat de normalité asymptotique pour $\hat{\lambda}$ pour déduire un intervalle de confiance asymptotique pour λ de coefficient de sécurité 95%. On donne $n = 100$ et $\bar{x} = 4,42$.

5. Soit $\ell \in \mathbb{N}^*$ fixé et on veut estimer la probabilité $p_\ell := e^{-\lambda} \frac{\lambda^\ell}{\ell!}$ d'avoir ℓ clients devant le guichet ł'heure d'observation. On note N_ℓ la variable aléatoire nombre d'observations (parmi les n) égales à ℓ.

a) Montrer que $N_\ell \sim \mathcal{B}(n, p_\ell)$. En déduire que $\frac{N_\ell}{n}$ est un estimateur sans biais de p_ℓ. Est-il convergent ?

b) On note $S = n\hat{\lambda}$ et $T = \mathrm{E}\left[\frac{N_\ell}{n} \mid S\right]$. Justifier les propriétés remarquables de l'estimateur T pour p_ℓ.

c) Quelles sont les lois de S et de $X_2 + \ldots + X_n$? Montrer que $T = \mathrm{P}(X_1 = \ell \mid S)$ et que

$$\mathrm{P}(X_1 = \ell \mid S = s) = \begin{cases} C_s^\ell \frac{(n-1)^{s-\ell}}{n^s} & \text{si } \ell \leqslant s \\ 0 & \text{si } \ell > s. \end{cases}$$

En déduire l'estimateur efficace de p_ℓ.

Exercice II. Une machine remplit des sachets avec un médicament. On voudrait avoir des informations sur la loi de la variable aléatoire poids d'un sachet. On prélève au hasard 21 sachets dans la production de cette machine dont les poids, exprimés en mg, sont : $0,3; 0,7; 0,9; 1,2; 1,4; 1,4; 1,5; 1,5; 1,6; 1,9; 2,0; 2,1; 2,1; 2,3; 2,5; 2,6; 2,7; 3,0; 3,8; 3,9; 4,0$. Tous les tests suivants seront construits au niveau $\alpha = 5\%$.

1) On suppose que la loi est $\mathcal{N}(m, 1)$. Tester l'hypothèse (H) : $m = 2$ contre (A) : $m = 3$ à l'aide d'un test de Neyman-Pearson. On donne $\bar{x} = 2.067$.

2) On suppose que la loi est $\mathcal{N}(m, \sigma^2)$, avec σ^2 inconnu. Utiliser le test de rapport de vraisemblance pour tester une fois de plus l'hypothèse (H) : $m = 2$. On donne $s^2 = 0,988$.

3) On ne connaît plus la loi. Expliquer et appliquer le test de Kolmogorov-Smirnov pour montrer que la loi de la variable aléatoire poids d'un sachet est $\mathcal{N}(2, 1)$. Dans les tables de Kolmogorov-Smirnov on trouve pour $n = 21$ la valeur $c = 0,2868$ pour $\alpha = 5\%$ et les calculs donnent $d_{21} = 0,1069$ (expliquer la signification et la façon dont on calcule en pratique cette valeur).

Exercice III. Deux échantillons indépendants $(X_1, \ldots, X_n)$ et $(Y_1, \ldots, Y_m)$ sont de lois respectivement $\gamma(1, \theta)$ et $\gamma(1, \theta')$. Quelles sont les lois de $U := 2\theta \sum_{i=1}^{n} X_i$ et de $V := 2\theta' \sum_{j=1}^{m} Y_j$ et de U/V. En déduire un intervalle de confiance pour $\rho := \frac{\theta}{\theta'}$, de coefficient de sécurité 90%. On donne $n = 6$, $m = 4$, $\bar{x} = 4,846$ et $\bar{y} = 3,323$.

Exercice IV. Soient $X_1, \ldots, X_n$ variables aléatoires indépendantes de même loi uniforme sur l'intervalle $[\theta, 2\theta]$, où le paramètre $\theta \in]0, \infty[$ est inconnu.

1) écrire la vraisemblance de l'échantillon. S'agit-il d'un modèle exponentiel ?

2) Trouver une statistique exhaustive et donner sa densité de probabilité.

3) On pose

$$U_1 := \frac{n+1}{2n+1} X_{(n)} \text{ et } U_2 := \frac{n+1}{5n+4}[2X_{(n)} + X_{(1)}]$$

Montrer que U_1 et U_2 sont deux estimateurs sans biais de θ. Lequel est meilleur ? On pourra d'abord montrer que $4\mathrm{Cov}(X_{(n)}, X_{(1)}) + \mathrm{Var}(X_{(1)}) < 0$ pour comparer U_1 et U_2.

Sujet d'examen partiel

Maîtrise de Mathématiques - Statistiques
samedi 27 mars 2004 - durée 3 heures - calculatrice interdite
documents (cours manuscrit ou polycopié) autorisés

Exercice I.

Zeus propose à Pythagore d'accéder à l'immortalité en jouant à pile ou face. Zeus, dieu suprême, justicier et bienveillant, sait que la pièce est truquée et qu'elle tombe sur pile avec probabilité $\theta \in]0,1[$, mais il dit seulement à Pythagore que la pièce est truquée. Pythagore lui répond qu'il accepte de jouer s'il le laisse d'abord lancer cette pièce n fois. En fait Pythagore veut se faire une idée sur la valeur de θ en utilisant les résultats $\mathbf{x} = (x_1, \ldots, x_n)$ des n lancers.

1) Proposer un modèle statistique pour cette expérience aléatoire. S'agit-il d'un modèle paramétrique ? écrire la densité de ce modèle par rapport à la mesure de comptage.

2) On note $\mathbf{X} = (X_1, \ldots, X_n)$ l'échantillon aléatoire dont l'observation est $\mathbf{x}$. Indiquer l'estimateur du maximum de vraisemblance $\hat{\theta}^*$ pour θ, ainsi qu'une statistique exhaustive complète S. En déduire un estimateur efficace pour θ. Est-il R-efficace (donner l'expression de l'information de Fisher) ?

3) Pythagore a l'habitude d'utiliser les carrés (sic !) et il veut estimer $q(\theta) = \theta^2$. Déduire l'expression de l'estimateur T_n du maximum de vraisemblance pour θ^2 à partir de $\hat{\theta}^*$. On pose $T_n := t_n(X_1, \ldots, X_n)$. Écrire l'expression de la fonction $t_n : \mathbb{R}^n \to \mathbb{R}$.

4) Pythagore constate que $T_n(X)$ est biaisé (le justifier !) ; cet estimateur n'est donc pas efficace. Il construit un autre estimateur pour θ^2 comme suit. Pour $i = 1, \ldots, n$, soit

$$T_n^{(i)} := t_{n-1}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$$

(autrement dit, on reprend le même estimateur qu'au point précédent mais on "oublie" le résultat du i-ème lancer). On note

$$J_n = nT_n - \frac{n-1}{n} \sum_{i=1}^{n} T_n^{(i)}.$$

Montrer que J_n est un estimateur sans biais de θ^2.

5) Prouver que $J_n = \dfrac{S(S-1)}{n(n-1)}$. On pourra utiliser dans le calcul les variables $S - X_i$, $i = 1, \ldots, n$.

6) J_n est-il efficace ? Si oui, le prouver ; sinon, construire un estimateur efficace à l'aide d'une statistique exhaustive pour θ^2.

7) Calculer l'expression de l'information de Fisher pour le paramètre θ^2 (éventuellement à partir de l'information de Fisher pour θ).

8) Soit $G_S(t) = \mathrm{E}_\theta\left[t^S\right]$, $t \in \mathbb{R}$, la fonction génératrice de S. Vérifier que :

$$\mathrm{E}_\theta\left[S(S-1)\right] = G_S''(1) \text{ et } \mathrm{E}_\theta\left[(S(S-1))^2\right] = \left[\frac{d^2}{dt^2}\left(t^2 G_S''(t)\right)\right]_{t=1}.$$

9) Montrer que $G_S(t) = (\theta t + 1 - \theta)^n$ et calculer $\mathrm{Var}_\theta(S(S-1))$.

10) En déduire l'expression de $\mathrm{Var}_\theta(J_n)$. étudier la R-efficacité de J_n.

épilogue : Pythagore reste immortel dans le souvenir des Hommes.

Exercice II.

La fiabilité (la durée de vie) d'un composant électronique dépend de deux caractères aléatoires indépendants X et Y, de lois exponentielles de paramètres respectifs $\lambda > 0$ et $\mu > 0$. Il n'est pas possible de mesurer directement les valeurs de X et Y, mais on peut mesurer avec un appareil les valeurs des variables aléatoires $Z = \min\{X, Y\}$ et $W = \begin{cases} 1 & \text{si } Z = X \\ 0 & \text{si } Z = Y \end{cases}$ (on dit que les données X et Y sont censurées).

1) Trouver la loi du couple (Z, W). On pourra calculer $\mathrm{P}(Z > t, W = \ell)$, pour tout $t \in \mathbb{R}$ et $\ell = 0, 1$.

2) En déduire que les lois de Z et W sont respectivement $\mathcal{E}(\lambda + \mu)$ et $\mathcal{B}(1, \dfrac{\lambda}{\lambda+\mu})$.

3) Prouver que les variables Z et W sont indépendantes.

On veut avoir des informations sur le paramètre $\theta = (\lambda, \mu)$ à partir d'un n-échantillon d'observations censurées $(Z_1, W_1), \ldots, (Z_n, W_n)$.

4) écrire la vraisemblance de cet échantillon.

5) Trouver le(s) point(s) $\hat{\theta}$ de maximum de la vraisemblance. En déduire l'estimateur du maximum de vraisemblance $\hat{\theta}^*$ pour θ. Montrer que $\hat{\theta}^* := (\hat{\lambda}^*, \hat{\mu}^*) = (\dfrac{\overline{W}}{\overline{Z}}, \dfrac{1-\overline{W}}{\overline{Z}})$, où $\overline{Z}$ et $\overline{W}$ sont les moyennes empiriques $\overline{Z}$ et $\overline{W}$ des vecteurs $\mathbf{Z} = (Z_1, \ldots, Z_n)$ et $\mathbf{W} = (W_1, \ldots, W_n)$, respectivement.

6) Les estimateurs $\hat{\lambda}^*$ et $\hat{\mu}^*$ sont-ils biaisés ? asymptotiquement sans biais ? efficaces ?

7) Prouver que chacun des estimateurs $\hat{\lambda}^*$ et $\hat{\mu}^*$ est consistant (sans utiliser le résultat de convergence pour l'estimateur du maximum de vraisemblance).

8) Justifier que, lorsque $n \to \infty$, $\left(\sqrt{n}(\overline{Z} - \dfrac{1}{\lambda+\mu}), \sqrt{n}(\overline{W} - \dfrac{\lambda}{\lambda+\mu})\right) \xrightarrow{\text{loi}} (G_1, G_2)$, où $G_1 \sim \mathcal{N}\left(0, \dfrac{1}{(\lambda+\mu)^2}\right)$ et $G_2 \sim \mathcal{N}\left(0, \dfrac{\lambda\mu}{(\lambda+\mu)^2}\right)$ sont indépendantes. On pourra calculer l'espérance et la matrice de covariance de chacun des vecteurs (Z_i, W_i), $i = 1, \ldots, n$.

9) Vérifier que $\sqrt{n}(\overline{W} - \lambda\overline{Z}) = \sqrt{n}(\overline{W} - \dfrac{\lambda}{\lambda+\mu}) - \lambda\left[\sqrt{n}(\overline{Z} - \dfrac{1}{\lambda+\mu})\right]$. En déduire que $\sqrt{n}(\overline{W} - \lambda\overline{Z})$ converge en loi vers $G_2 - \lambda G_1$. On pourra utiliser la définition de la convergence en loi ainsi que l'application $(a, b) \mapsto b - \lambda a$. Quelle est la loi de $G_2 - \lambda G_1$?

10) En déduire que $\sqrt{n}(\hat{\lambda}^* - \lambda)$ converge en loi vers une variable gaussienne centrée dont on calculera la variance. Que peut-on dire de la suite $\sqrt{n}(\hat{\mu}^* - \mu)$?

Sujet d'examen

Maîtrise de Mathématiques - Statistiques
mercredi 9 juin 2004 - durée 3 heures
documents (cours manuscrit ou polycopié) et calculatrice autorisés

Exercice I. Un statisticien connaît l'allure du graphe d'une densité f sur $\mathbb{R}$, mais son expression est inconnue. Il observe $(X_1, \ldots, X_n)$ un n-échantillon de densité f et il considère le problème de test non-paramétrique $\mathrm{H}_0 : f = f_0$ contre $\mathrm{H}_1 : f = f_1$. Ici f_0 et f_1 sont deux densités définies sur $\mathbb{R}$ dont il connaît les expressions et qui n'ont pas de paramètre inconnu.

1. Montrer que pour tout $\lambda \in [0, 1]$ la fonction $(1 - \lambda)f_0 + \lambda f_1$ est une densité de probabilité. Réduire le problème de test non-paramétrique à un problème de test classique (les hypothèses H_0 et H_1 s'expriment à l'aide d'un paramètre).

2. Soient $f_0(x) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{1}{2}x^2)$ et $f_1(x) = \frac{1}{2} \exp(-|x|)$, $x \in \mathbb{R}$.

a) Montrer que le meilleur test (dans un sens à préciser) est basé sur une région de rejet de la forme $\sum_{i=1}^{n}(|x_i| - 1)^2 \geqslant c$ avec c une constante.

b) Calculer les erreurs de première et seconde espèce lorsqu'on possède une seule observation ($n = 1$) et pour $c = 1$ ainsi que pour $c = \frac{1}{4}$. Le test est-il sans biais ?

c) Montrer qu'on peut bâtir un test de niveau $\alpha = 0,05$ basé sur une seule observation en choisissant une valeur particulière de c. A-t-on unicité ? Expliquer la décision si on observe $x_1 = 1,83$.

Exercice II. On observe $(X_1, \ldots, X_n)$ un n-échantillon de loi géométrique $\mathcal{G}(\frac{1}{\theta})$, $\theta > 1$.

1. Montrer qu'il s'agit d'une loi de type exponentiel. En déduire une statistique suffisante complète. Que vaut l'information de Fisher $I(\theta)$?

2. Trouver un estimateur par la méthode des moments pour θ et montrer qu'il coïncide avec l'estimateur du maximum de vraisemblance pour θ. Cet estimateur est-il R-efficace ? efficace ? Que peut-on dire quant à sa convergence et sa normalité asymptotique ?

3. En déduire un intervalle de confiance asymptotique pour θ de coefficient de sécurité $0,90$.

Exercice III.

1. Une variable aléatoire ε est de loi inconnue, mais on soupçonne que la loi est $\mathcal{N}(0, 1)$.

a) Soient $V_1 =]-\infty, a]$, $V_2 =]a, b]$, $V_3 =]b, c]$ et $V_4 =]c, \infty[$ tels que $\mathrm{P}(G \in V_j) = \frac{1}{4}$, $j = 1, .., 4$, où $G \sim \mathcal{N}(0, 1)$. Trouver les valeurs de a, b et c.

b) On observe un 31-échantillon de ε et on trouve les valeurs : $-1,7$; $-1,5$; $-1,3$; $-1,1$; $-0,8$; $-0,6$; $-0,6$; $-0,62$; $-0,5$; $-0,5$; $-0,41$; $-0,24$; $-0,01$; $-0,01$; 0; $0,02$; $0,1$; $0,1$; $0,26$; $0,29$; $0,3$; $0,5$; $0,6$; $0,7$; $0,93$; 1; $1,2$; $1,75$; $1,8$; $1,9$; 2.

Calculer le nombre n_j de valeurs se trouvant dans V_j, $j = 1, .., 4$. Appliquer un test de χ^2 au niveau 0,05 pour voir si l'échantillon est bien de loi gaussienne standard.

2. On accepte l'hypothèse que $(\varepsilon_1, \ldots, \varepsilon_n)$ est un n-échantillon de loi $\mathcal{N}(0,1)$. Supposons que les variables aléatoires $Y_1, \ldots, Y_n$ vérifient $Y_i = \beta x_i + \varepsilon_i$, $i = 1, \ldots, n$, où $x_1, \ldots, x_n$ sont des constantes non nulles fixées et β est un paramètre réel. Pour estimer β on considère les estimateurs suivants : $\beta_1^* = [\sum_{i=1}^n (Y_i/x_i)]/n$, $\beta_2^* = (\sum_{i=1}^n Y_i)/(\sum_{i=1}^n x_i)$ et enfin $\hat{\beta}^*$ l'estimateur du maximum de vraisemblance. Montrer ces trois estimateurs sont sans biais et comparer leurs variances. Lequel est le meilleur ?

3. Dans la suite on supposera que la loi commune des ε_i est $\mathcal{N}(0, \sigma^2)$ avec $\sigma^2 > 0$ inconnu.

a) On note $S_{xx} = \sum_{i=1}^n x_i^2$ et $S_{xY} = \sum_{i=1}^n x_i Y_i$. Montrer que l'estimateur de maximum de vraisemblance de (β, σ^2) est $(\hat{\beta}^*, \widehat{\sigma^2}^*) = \left(\frac{S_{xY}}{S_{xx}}, \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\beta}^* x_i)^2\right)$.

b) Montrer que $\hat{\beta}^*$ est un estimateur sans biais de β et calculer sa variance.
On pose $\hat{\varepsilon}_i := Y_i - \hat{\beta}^* x_i$. Montrer que $\mathrm{E}(\hat{\varepsilon}_i) = 0$ et $\mathrm{Cov}(Y_i, \hat{\beta}^*) = \sigma^2 x_i / S_{xx}$.
En déduire $\mathrm{Var}(\hat{\varepsilon}_i)$ et prouver ensuite que $\widehat{\sigma^2}^*$ est un estimateur biaisé de σ^2.
Soit S^2 l'estimateur sans biais de σ^2 obtenu à partir de $\widehat{\sigma^2}^*$.

c) Montrer que $\hat{\beta}^* \sim \mathcal{N}(\beta, \sigma^2/S_{xx})$.
On veut utiliser le théorème de Cochran pour déduire que les variables aléatoires $\hat{\beta}^*$ et S^2 sont indépendantes et que $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.
On note $\mathbf{x}$, $\mathbf{Y}$ et ε les vecteurs de coordonnées respectivement x_i, Y_i et ε_i, ainsi que $V := \mathrm{Vect}(\mathbf{x})$. Montrer que $\frac{1}{\sigma}(\hat{\beta}^* - \beta) = \frac{<\delta, \mathbf{x}>}{\|\mathbf{x}\|^2}$ et que $\frac{(n-1)S^2}{\sigma^2} = \|\delta - \frac{<\delta,\mathbf{x}>}{\|\mathbf{x}\|^2}\mathbf{x}\|^2$, où $\delta = \frac{1}{\sigma}\varepsilon$ est un vecteur gaussien dont on précisera la loi. Conclure.

d) Prouver que $\frac{\hat{\beta}^* - \beta}{\sqrt{S^2/S_{xx}}}$ suit une loi de Student$(n-1)$. En déduire un intervalle de confiance de coefficient de sécurité $1 - \alpha$ pour β. Utiliser cet intervalle pour bâtir un test de niveau α pour tester $\beta = 0$ contre $\beta \neq 0$.
Application : $\alpha = 0,05$, $x_1 = 5$, $x_2 = 10$, $x_3 = 15$, $x_4 = 20$, $x_5 = 25$, $x_6 = 30$ et $y_1 = 0,1$; $y_2 = 0,21$; $y_3 = 0,3$; $y_4 = 0,38$; $y_5 = 0,49$; $y_6 = 0,62$.

e) Supposons qu'il existe $\lim_{n\to\infty} \frac{1}{n} S_{xx} = \ell$, une constante. On veut montrer que $\hat{\beta}^*$ et $\widehat{\sigma^2}^*$ sont des estimateurs consistants.
Calculer les fonctions caractéristiques de $x_i Y_i$ et de S_{xY}. En déduire que $\frac{1}{n} S_{xY}$ converge en loi et en probabilité vers $\beta\ell$, quand $n \to \infty$. Calculer la limite en probabilité de $\hat{\beta}^*$.
Quelle est la limite en probabilité de $\frac{1}{n}\sum_{i=1}^n (Y_i - \beta x_i)^2$? En déduire que $\frac{1}{n}\sum_{i=1}^n Y_i^2$ converge en probabilité vers $\sigma^2 + \beta^2\ell$, quand $n \to \infty$. Calculer ensuite la limite en probabilité de $\widehat{\sigma^2}^*$.

Sujet d'examen de septembre

Maîtrise de Mathématiques - Statistiques
vendredi 3 septembre 2004 - durée 3 heures
documents (cours manuscrit ou polycopié) et calculatrice autorisés

Exercice I. On observe un n-échantillon de loi normale $\mathcal{N}(\theta_1, \theta_2)$, où $(\theta_1, \theta_2) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$.

1. Montrer que la vraisemblance de l'échantillon admet une écriture de type exponentiel et indiquer les quantités apparaissant dans cette écriture. Indiquer une statistique exhaustive complète. En déduire un estimateur sans biais pour le paramètre (θ_1, θ_2).

2. On suppose dans la suite que $\theta_1 = \theta > 0$ et que $\theta_2 = \theta^2$.

a) Montrer que le couple $(\overline{X}, S_0^2) = (\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2)$ est une statistique exhaustive pour θ. Est-elle complète ?

b) Vérifier que $\overline{X}$ est un estimateur sans biais pour θ et écrire ses propriétés de convergence. Trouver la constante c telle que $c\,S_0$ soit aussi un estimateur sans biais pour θ, avec $S_0 = \sqrt{S_0^2}$. Est-il consistant ?

c) Montrer que pour tout nombre $b \in \mathbb{R}$, l'estimateur $b\overline{X} + (1-b)(c\,S_0)$ est un estimateur sans biais de θ. Déduire la valeur de b pour laquelle on trouve un estimateur T_1 sans biais de variance minimale pour θ.

d) Soit $b_1\overline{X} + b_2(c\,S_0)$ un autre estimateur pour θ tel que $b_1 + b_2 \neq 1$. Trouver les nombres $b_1, b_2 \in \mathbb{R}$ tels que la quantité $\mathrm{E}_\theta\left[(\theta - b_1\overline{X} - b_2(c\,S_0))^2\right]$ soit la plus petite possible. Notons par T_2 l'estimateur de la forme $b_1\overline{X} + b_2(c\,S_0)$ obtenu avec ces valeurs. Comparer les qualités quadratiques $d_1^2(\theta)$ et $d_2^2(\theta)$ des deux estimateurs T_1 et T_2 et indiquer le meilleur.

3. On continue de supposer que la loi commune des X_i est $\mathcal{N}(\theta, \theta^2)$.

a) Calculer l'information de Fisher $I(\theta)$. Étudier la R-efficacité et l'efficacité de chacun des estimateurs $\overline{X}$ et $c\,S_0$.

b) Utiliser les estimateurs $\overline{X}$ de θ et S_0^2 de θ^2 pour construire des intervalles de confiance de coefficient de securité $1-\alpha \in]0,1[$ pour θ.

4. On suppose cette fois-ci que la loi commune des X_i est $\mathcal{N}(\theta, a\theta^2)$, où a est un paramètre positif inconnu. On veut tester l'hypothèse (H) : $a = 1$ contre (A) : $a \neq 1$. θ est ici un paramètre de nuisance inconnu. Construire le test de rapport de vraisemblance de niveau $\alpha \in]0,1[$.

Exercice II. On possède une seule observation X_1 d'un caractère de loi donnée par

$$\mathrm{P}(X_1 = -1) = \frac{\theta}{2},\ \mathrm{P}(X_1 = 0) = 1 - \theta,\ \mathrm{P}(X_1 = 1) = \frac{\theta}{2},\ 0 \leqslant \theta \leqslant 1.$$

1. écrire la densité de X_1 par rapport à la mesure de comptage. S'agit-il d'une loi de type exponentiel ?

2. X_1 est-elle une statistique suffisante complète ? Même question pour $|X_1|$.

3. Trouver l'estimateur du maximum de vraisemblance de θ. Est-il sans biais ?

4. Montrer que $T(X_1) = \begin{cases} 2 & \text{si } X_1 = 1 \\ 0 & \text{sinon} \end{cases}$ est un estimateur sans biais pour θ.

5. Trouver un meilleur estimateur que $T(X_1)$ et prouver qu'il est meilleur.

Exercice III. Une variable aléatoire X prend seulement les valeurs $\{1, \dots, 7\}$ avec des probabilités différentes sous deux hypothèses (H) et (A). Ainsi :

$$\text{sous (H)} \quad X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0,01 & 0,01 & 0,01 & 0,01 & 0,01 & 0,01 & 0,94 \end{pmatrix}$$

et

$$\text{sous (A)} \quad X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0,06 & 0,05 & 0,04 & 0,03 & 0,02 & 0,01 & 0,79 \end{pmatrix}.$$

Utiliser le lemme de Neyman-Pearson pour construire le test pur le plus puissant pour (H) contre (A) de niveau $\alpha = 0,04$ basé sur une seule observation de X. Calculer la puissance de ce test.

Exercice IV. Une espèce de tulipes peut avoir trois couleurs : rouge, rose et violet. Un statisticien élabore un modèle : il pense que ces couleurs apparaissent avec probabilités, respectivement $\frac{2}{3}$, $\frac{1}{4}$ et $\frac{1}{12}$. Il commande 60 tulipes au hasard et il reçoit 36 rouges, 18 roses et 6 violets. Tester la validité du modèle au niveau $\alpha = 0,05$.

Sujet d'examen partiel

Maîtrise de Mathématiques et IUP GMI 3ème année - Statistiques
vendredi 1er avril 2005 - durée 3 heures - calculatrice interdite
documents (cours manuscrit ou polycopié) autorisés

Exercice I.

On observe le nombre de clients devant un guichet de banque, à la même heure, durant n jours choisis d'une manière indépendante $\mathbf{x} := (x_1, \ldots, x_n)$. On admet que la loi qui modélise ce phénomène aléatoire est une loi de Poisson.

1. Construire effectivement le modèle statistique associé à cette expérience. On ne connaît pas le paramètre $\lambda > 0$ de la loi de Poisson. à partir d'un n-échantillon $\mathbf{X}$ indiquer deux estimateurs de λ par la méthode des moments. Sont-ils convergents ?

2. Montrer que parmi les deux estimateurs précédents celui qui est sans biais est aussi l'estimateur du maximum de vraisemblance $\hat{\lambda}^*$. S'agit-il d'un estimateur exhaustif ? complet ? R-efficace ? efficace ?

3. Que vaut la limite en loi de $\sqrt{n}(\hat{\lambda}^* - \lambda)$? Trouver une fonction borélienne $g : \mathbb{R}_+^* \to \mathbb{R}_+^*$ telle que la limite en loi de la suite de variables aléatoires $\sqrt{n}(g(\hat{\lambda}^*) - g(\lambda))$ soit indépendante de λ.

4. Soit $\ell \in \mathbb{N}^*$ fixé et on veut estimer la probabilité $p_\ell := e^{-\lambda}\frac{\lambda^\ell}{\ell!}$ d'avoir ℓ clients devant le guichet à l'heure d'observation. On note $N_\ell = \sum_{i=1}^n \mathbb{1}_{\{X_i = \ell\}}$ la variable aléatoire nombre d'observations égales à ℓ. Montrer que $N_\ell \sim \mathcal{B}(n, p_\ell)$. En déduire que $\frac{N_\ell}{n}$ est un estimateur sans biais de p_ℓ. Est-il convergent ?

5. On note $S := n\hat{\lambda}^*$ et $T := \mathrm{E}\left[\frac{N_\ell}{n} \mid S\right]$. Donner les propriétés de T en tant qu'estimateur de p_ℓ.

6. Montrer que $T = \mathrm{P}(X_1 = \ell \mid S)$. Quelles sont les lois de S et de $X_2 + \ldots + X_n$? Montrer que $\mathrm{P}(X_1 = \ell \mid S = s) = \begin{cases} C_s^\ell \frac{(n-1)^{s-\ell}}{n^s} & \text{si } \ell \leqslant s \\ 0 & \text{si } \ell > s \end{cases}$. Déduire l'estimateur efficace de p_ℓ.

Exercice II.

Le prix en bourse d'une option européenne est une variable aléatoire X_1 de densité

$$f_{a,b}(x_1) = (2\pi b x_1^2)^{-1/2} \exp\left[-(\ln x_1 - a)^2/(2b)\right] \mathbb{1}_{x_1 > 0},$$

où les paramètres $a \in \mathbb{R}$ et $b > 0$ sont inconnus. Pour faire de la prévision il faut préciser mieux cette loi en utilisant l'observation d'un n-échantillon $\mathbf{X} = (X_1, \ldots, X_n)$. Un trader utilise une méthode "ad-hoc", en appliquant un "filtre" logarithmique : il pose $Y_i := \ln X_i$, $i = 1, \ldots, n$. Ensuite, à partir de l'échantillon $\mathbf{Y} = (Y_1, \ldots, Y_n)$, il donne une estimation du paramètre $\theta = (a, b)$. Un étudiant ayant fait des statistiques utilise plutôt $m = \mathrm{E}(X_1)$ et $\sigma^2 = \mathrm{Var}(X_1)$ pour estimer le paramètre $\lambda = (m, \sigma^2)$, à partir de l'échantillon $\mathbf{X}$.

1) Trouver la densité de Y_1 et calculer $\mathrm{E}(e^{\alpha Y_1})$, $\alpha \in \mathbb{R}$.

2) Le trader utilise l'estimateur $\theta^* = (\overline{Y}, S_Y^2)$. Par quelle(s) méthode(s) peut-on l'obtenir ? Est-il R-efficace ? efficace ? Est-il consistant ? asymptotiquement normal ? Que vaut sa matrice de covariance asymptotique, notée $I(\theta)^{-1}$? On pourra utiliser les résultats de cours sans faire des calculs.

3) Utiliser le résultat de **1)** pour calculer λ en fonction de θ. On notera dans la suite $\lambda = q(\theta)$, $q : \mathbb{R} \times \mathbb{R}_+^* \to \mathbb{R}_+^* \times \mathbb{R}_+^*$.

4) Écrire la vraisemblance de l'échantillon $\mathbf{X}$ et indiquer une statistique exhaustive complète pour θ ainsi que pour λ. Pourquoi θ^* est une statistique exhaustive complète pour λ ?

5) Trouver l'estimateur du maximum de vraisemblance $\hat{\lambda}^*$ pour λ. On pourra éventuellement utiliser l'estimateur du maximum de vraisemblance pour θ.

6) Énoncer le résultat de normalité asymptotique pour $\hat{\lambda}^*$. Calculer la matrice jacobienne $\mathrm{Jac}_q(\theta) = \big(\nabla m(\theta) \quad \nabla\sigma^2(\theta)\big)$. Montrer (ou éventuellement admettre) que la matrice de covariance asymptotique de $\hat{\lambda}^*$ est donnée par $\mathrm{Jac}_q(\theta)^\star I(\theta)^{-1}\mathrm{Jac}_q(\theta)$ et la calculer. En déduire que la borne de Cramer-Rao pour le paramètre m seul est $\kappa := \frac{m^2}{n}b(1+\frac{b}{2})$ (on pourra utiliser un élément de la matrice de covariance asymptotique).

7) Dans la suite on étudiera l'estimation de seul m (σ^2 sera un paramètre de nuisance inconnu). Montrer que $m_0^* := \overline{X}$ et $\widehat{m}^* := \exp(\overline{Y} + S_Y^2/2)$ sont deux estimateurs par la méthode des moments de m, les deux étant consistants. Calculer $\mathrm{Var}_{a,b}(m_0^*)$ et la comparer avec κ, la borne de Cramer-Rao. Calculer $\mathrm{E}_{a,b}[\widehat{m}^*]$ (on utilisera les propriétés de $\overline{Y}$ et S_Y^2). Montrer que $\mathrm{E}_{a,b}[\widehat{m}^*] > m$, mais que $\lim_{n\to\infty} \mathrm{E}_{a,b}[\widehat{m}^*] = m$ Commenter vos résultats.

8) Calculer $\mathrm{E}_{a,b}[(\widehat{m}^*)^2]$. Montrer que, lorsque $n \to \infty$, $\mathrm{E}_{a,b}[(\widehat{m}^*)^2] = e^{2a+b}[1+\frac{b+b^2}{n}+o(\frac{1}{n})]$ et $\mathrm{E}_{a,b}[\widehat{m}^*]^2 = e^{2a+b}[1 + \frac{b^2}{2n} + o(\frac{1}{n})]$. On étudiera d'abord les développements limités des logarithmes des quantités en question. En déduire le développement limité de $\mathrm{Var}_{a,b}(\widehat{m}^*)$ ainsi que $\lim_{n\to\infty} \frac{1}{\kappa}\mathrm{Var}_{a,b}(\widehat{m}^*)$? Commenter ce résultat.

9) L'étudiant pense que le biais de $\widehat{m}^*$ provient du facteur $\exp(S_Y^2/2)$. Justifier que si ϕ est une fonction telle que $\mathrm{E}_{a,b}[\phi(S_Y^2)] = \exp[(b/2)(1-1/n)]$, alors l'estimateur $m_1^* := \exp(\overline{Y})\phi(S_Y^2)$ est efficace pour m. On veut trouver une telle fonction ϕ. Développer en série entière de b la fonction $\exp[(b/2)(1-1/n)]$. Montrer que, pour tout $\ell \in \mathbb{N}$, $\mathrm{E}_{a,b}[(nS_Y^2)^\ell] = 2^\ell b^\ell \frac{\Gamma(\ell+(n-1)/2)}{\Gamma((n-1)/2)}$. Déduire que $\phi(S_Y^2)$ qui répond aux exigences requises est $\phi(S_Y^2) = \Gamma(\frac{n-1}{2}) \sum_{\ell=0}^\infty \left(\frac{n-1}{2}\right)^\ell \frac{1}{\ell!} \frac{1}{\Gamma(\ell+(n-1)/2)} (S_Y^2/2)^\ell$.

10) Notons $U_n := e^{\overline{Y}}$, $V_n := \sqrt{n}(e^{\overline{Y}} - e^a)$ et $W_n := \sqrt{n}(e^{S_Y^2/2} - e^{b/2})$. Montrer que, lorsque $n \to \infty$, $U_n \to e^a$ en probabilité et que $V_n \to e^a\sqrt{b}G_1$ en loi et $W_n \to e^{b/2}\frac{b}{\sqrt{2}}G_2$ en loi, avec $G_1, G_2 \sim \mathcal{N}(0,1)$ deux variables aléatoires indépendantes. On utilisera la normalité asymptotique de θ^* obtenue au point **2)**. En déduire que $(U_n, V_n) \to (e^a, e^a\sqrt{b}G_1)$ en loi et ensuite que $(U_n, V_n, W_n) \to (e^a, e^a\sqrt{b}G_1, e^{b/2}\frac{b}{\sqrt{2}}G_2)$ en loi. Soit la fonction continue $\psi : \mathbb{R}^3 \to \mathbb{R}$, $\psi(u,v,w) = uw + e^{b/2}v$. Montrer que $\psi(U_n, V_n, W_n) = \sqrt{n}(\widehat{m}^* - m)$. En déduire la convergence en loi de $\sqrt{n}(\widehat{m}^* - m)$ et retrouver sa limite en loi.

Sujet d'examen

Maîtrise de Mathématiques - Statistiques
mardi 31 mai 2005 - durée 3 heures
documents (cours manuscrit ou polycopié) et calculatrice autorisés

Exercice I. Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un échantillon de densité $f_\theta(z) = \begin{cases} ce^{-z/\theta_1}, & \text{si } z > 0 \\ ce^{z/\theta_2}, & \text{si } z < 0 \end{cases}$, où $\theta = (\theta_1, \theta_2) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ et $c > 0$ est une constante.

1. Que vaut la constante c ? Montrer que $(S_1, S_2) = (\sum_{i=1}^n |X_i| \mathbb{1}_{\{X_i>0\}}, \sum_{i=1}^n |X_i| \mathbb{1}_{\{X_i<0\}})$ est une statistique exhaustive pour θ. Calculer $\mathrm{E}(S_1)$ et $\mathrm{E}(S_2)$.
2. Trouver l'estimateur du maximum de vraisemblance $\hat{\theta}^*$ de θ en fonction de S_1 et S_2. Étudier la consistance de $\hat{\theta}^*$.

Exercice II. On note $X_i = (\theta/2)t_i^2 + \varepsilon_i$, $i = 1, \ldots, n$, où ε_i sont des variables aléatoires indépendantes de même loi $\mathcal{N}(0, 1)$ et $\theta \in \mathbb{R}$ est inconnu. On fixe $0 < \alpha < 1$.

1. écrire la vraisemblance du modèle et trouver l'estimateur du maximum de vraisemblance $\hat{\theta}^*$ de θ. Trouver la loi de $\hat{\theta}^*$ et l'utiliser pour trouver un intervalle de confiance de coefficient de sécurité $(1 - \alpha)$ pour θ.
2. Supposons que $0 \leqslant t_i \leqslant 1$, $i = 1, \ldots, n$, mais qu'on peut choisir librement les valeurs des t_i. Quelles valeurs doit-on prendre pour les t_i pour rendre l'intervalle du point précédent le plus court possible (α reste fixé).

Exercice III. On fait une seule observation de la variable aléatoire à valeurs entières L. Notons f sa densité par rapport à la mesure de comptage. On veut tester (H) : $f = f_0$ contre (A) : $f = f_1$. Ici f_0 est la densité de la loi $\mathcal{P}(1)$ et f_1 celle de la loi de $T - 1$, avec $T \sim \mathcal{G}(1/2)$.

1. Montrer que la région de rejet est donnée par $\{\ell \in \mathbb{N} : (1,36)(\ell!/2^\ell) \geqslant c\}$.
2. Quand $c = 3$ trouver le niveau α du test basé sur cette région de rejet (on pourra vérifier que la suite $\{\ell!/2^\ell : \ell \in \mathbb{N}^*\}$ est croissante). Que vaut la puissance de ce test ?

Exercice IV. Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes ayant la même densité $f_\theta(z) = (1/b) \exp(-(z - a)/b) \mathbb{1}_{\{z \geqslant b\}}$, où $\theta = (a, b) \in \mathbb{R} \times \mathbb{R}_+^*$ (on notera cette loi $\mathcal{E}(a, b)$).

1. Soient $X_{(1)} \leqslant \ldots \leqslant X_{(n)}$ les statistiques d'ordre associés à l'échantillon et on pose $Y_i = (n - i + 1)[X_{(i)} - X_{(i-1)}]$, $i = 1, \ldots, n$ (ici $X_{(0)} = 0$). Montrer que la densité de $\mathbf{Y} = (Y_1, \ldots, Y_n)$ est : $\mathbf{y} \mapsto (1/b^n) \exp[-(1/b)(\sum_{i=1}^n y_i - na)] \mathbb{1}_{\{y_1 \geqslant na, y_2 \geqslant 0, \ldots, y_n \geqslant 0\}}$. En déduire la densité du vecteur aléatoire $(Y_1, U_1, \ldots, U_{n-2}, T_{n-1})$, où on a noté $U_1 = Y_2, U_2 = Y_2 + Y_3, \ldots, U_{n-2} = Y_2 + \ldots + Y_{n-1}$ et $T_{n-1} = Y_2 + \ldots + Y_n$. Montrer que la densité de (Y_1, T_{n-1}) est : $(y, t) \mapsto (1/b) \exp(-(y - na)/b) \mathbb{1}_{\{y \geqslant na\}} \cdot (t^{n-2}/(n-2)!b^{n-1}) \exp(-t/b) \mathbb{1}_{\{t \geqslant 0\}}$.

2. Montrer que :
• $S_1 := X_{(1)} \sim \mathcal{E}(a, b/n)$ est indépendante de $S_2 := \sum_{i=1}^n [X_i - X_{(1)}] \sim (b/2)\chi^2(2n - 2)$;
• (S_1, S_2) est une statistique suffisante pour θ (utiliser la définition de l'exhaustivité) ;
• (S_1, S_2) est complète ; c'est-à-dire que $\mathrm{E}_\theta[h(S_1, S_2)] = 0, \forall\theta \Rightarrow h = 0$ p.p. sur $\mathbb{R}^2$ (on pourra commencer par fixer $b \in \mathbb{R}^*$ et noter $\hbar(s_1, b) := \mathrm{E}_b[h(s_1, S_2)]$; montrer que pour p.t. $s_1 \in \mathbb{R}$, $\hbar(s_1, b) = 0$ p.p. en b, ensuite, par continuité, partout en $b \in \mathbb{R}$; conclure) ;
• $S_2/(n-1)$ et resp. $S_1 - (S_2/n(n-1))$ sont des estimateurs efficaces de b et resp. a ;
• lorsque $n \to \infty$, $n(S_1 - a)/b \to \mathcal{E}(0, 1)$ et $\sqrt{n}((S_2/n) - b) \to \mathcal{N}(0, b^2)$, en loi.

Exercice V. Soient $(X_1, \ldots, X_{n_1})$ et $(Y_1, \ldots, Y_{n_2})$ deux échantillons indépendants de lois respectivement $\mathcal{E}(\theta)$ et $\mathcal{E}(\lambda)$, $\theta > 0, \lambda > 0$. On pose $\Delta = \theta/\lambda$.

1. Montrer que $\Delta(\overline{X}/\overline{Y})$ est une fonction pivotale en lui trouvant sa loi. Soit $0 < \alpha < 1$ et notons par $f(\alpha)$ le réel positif satisfaisant $\mathrm{P}(F > f(\alpha)) = \alpha$, où F est de loi de Fisher de paramètre $(2n_1, 2n_2)$. Trouver $a > 0$ et $b > 0$ tels que $[a\overline{y}/\overline{x}, b\overline{y}/\overline{x}]$ soit un intervalle de confiance pour Δ de coefficient de sécurité $1 - \alpha$ (on pourra utiliser la fonction pivotale et des quantités de type $f(\alpha)$).

2. Montrer que le test ayant la région d'acceptation $\{f(1 - \alpha/2) \leqslant \overline{x}/\overline{y} \leqslant f(\alpha/2)\}$ est de niveau α pour tester (H) : $\Delta = 1$ contre (A) : $\Delta \neq 1$. Expliquer vos arguments. Peut-on obtenir une région d'acceptation de même type pour le test du rapport de vraisemblance (on pourra utiliser $t(\mathbf{x}, \mathbf{y}) = n_1\overline{x}/(n_1\overline{x} + n_2\overline{y})$ et les variations de la fonction $[0, 1] \ni t \mapsto t^{n_1}(1 - t)^{n_2}$).

Application : Une hydrocentrale utilise deux types de ventilateurs d'air dans la salle des turbines. Dans le tableau ci-dessous on donne les intervalles (en jours) entre les pannes survenues (la réparation est effectuée le jour même de la panne). Par expérience l'hypothèse de lois exponentielles est raisonnable. Donner un intervalle de confiance de coefficient de sécurité 90% pour le rapport Δ des temps moyens de fonctionnement. L'hypothèse (H) : $\Delta = 1$ est-elle rejetée au niveau $\alpha = 0,10$?

x	8	26	10	8	29	26														
y	3	150	40	34	32	37	34	2	31	6	5	14	150	27	4	6	27	10	30	37

Pour un seul de ces échantillons au choix vérifier la véracité de l'hypothèse de loi exponentielle par un test non-paramétrique (dont vous avez encore le choix). Expliquer votre démarche. On pourra considérer $n = 20$ comme “grand”.

Sujet d'examen de rattrapage

IUP GMI 3ème année - Statistiques
lundi 20 juin 2005 - durée 2 heures - calculatrice interdite
documents (cours manuscrit ou polycopié) autorisés

Exercice I. Les durées de vie de trois appareils identiques constituent un 3-échantillon X_1, X_2, X_3 de loi de fonction caractéristique $(1 - it\theta)^{-1}$.

1. De quelle loi remarquable s'agit-il ? Indiquer son espérance et sa variance.

2. Pourquoi X_2 est un estimateur sans biais de θ ? Calculer sa variance.

3. Montrer que $S = X_1 + X_2 + X_3$ est une statistique exhaustive pour θ. Est-elle complète ? minimale ?

4. Montrer que $\mathrm{E}_\theta(X_1 \mid S) = \mathrm{E}_\theta(X_2 \mid S) = \mathrm{E}_\theta(X_3 \mid S)$. On pourra d'abord calculer la densité du couple $f_{(X_1,S)}$ et ensuite déduire la densité conditionnelle $f_{X_1|S}$. Prouver que la valeur commune de ces espérances conditionnelles est $\frac{S}{3}$.

5. Indiquer un estimateur efficace θ^* pour θ. Justifier pourquoi X_2 n'est pas un estimateur efficace. Comparer les estimateurs θ^* et X_2. θ^* est-il R-efficace ?

Exercice II. Le taux de cholestérol dans le sang est un caractère aléatoire dont on peut supposer qu'il est de loi normale $\mathcal{N}(\theta, \theta^2)$, où $\theta \in \mathbb{R}_+^*$ est un paramètre inconnu. On observe un n-échantillon $X_1, \ldots, X_n$ ayant cette loi.

1. Montrer que la vraisemblance de l'échantillon admet une écriture de type exponentiel et indiquer les quantités apparaissant dans cette écriture. Montrer que le couple $(\overline{X}, S_0^2) = (\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2)$ est une statistique exhaustive pour θ. Est-elle complète ?

2. Vérifier que $\overline{X}$ est un estimateur sans biais pour θ et écrire ses propriétés de convergence. On note $S_0 = \sqrt{S_0^2}$. Trouver la constante c telle que $c\,S_0$ soit aussi un estimateur sans biais pour θ. Est-il consistant ?

3. Calculer l'information de Fisher $I(\theta)$. Étudier la R-efficacité et l'efficacité de chacun des estimateurs $\overline{X}$ et $c\,S_0$, où c est la constante trouvée au point précédent.

4. Avec la même valeur de la constante c, montrer que pour tout nombre $b \in \mathbb{R}$, l'estimateur $b\overline{X} + (1-b)(c\,S_0)$ est un estimateur sans biais de θ. Déduire la valeur de b pour laquelle on trouve un estimateur T_1 sans biais de variance minimale pour θ.

Sujet d'examen de rattrapage

Maîtrise de Mathématiques - Statistiques
vendredi 1er juillet 2005 - durée 2 heures
documents (cours manuscrit ou polycopié) et calculatrice autorisés

Exercice I. Soit X_1, X_2, X_3 un 3-échantillon de loi de fonction caractéristique $(1-it\theta)^{-1}$.

1. Calculer, à partir de la fonction caractéristique, l'espérance et la variance de cette loi.
2. Pourquoi X_2 est un estimateur sans biais de θ ? Calculer sa variance.
3. Montrer que $S = X_1 + X_2 + X_3$ est une statistique exhaustive pour θ. Est-elle complète ? minimale ?
4. Montrer que $\mathrm{E}_\theta(X_1 \mid S) = \mathrm{E}_\theta(X_2 \mid S) = \mathrm{E}_\theta(X_3 \mid S)$. On pourra d'abord calculer la densité du couple $f_{(X_1,S)}$ et ensuite déduire la densité conditionnelle $f_{X_1|S}$. Prouver que la valeur commune de ces espérances conditionnelles est $\frac{S}{3}$.
5. Indiquer un estimateur efficace θ^* pour θ. Justifier pourquoi X_2 n'est pas un estimateur efficace. Comparer les estimateurs θ^* et X_2. θ^* est-il R-efficace ?
6. Trouver la loi de la fonction pivotale $T_3(\theta) = \frac{2}{\theta}S$. En déduire qu'il existe deux réels $0 < a < b$ tels que $\mathrm{P}(a \leqslant \eta \leqslant b) = 0,95$, où η est une variable aléatoire ayant la même loi que $T_3(\theta)$. Si $a = 1,237$ quelle sera la valeur pour b ? En déduire un intervalle de confiance pour θ de coefficient de sécurité 0,95.
7. Trouver le test u.p.p. pour tester $\theta \geqslant \theta_0 = 1000$ contre $\theta < 1000$ au niveau 5%.
8. Quelle taille doit avoir l'échantillon telle que la puissance au point $\theta_1 = 105,5$ soit 95% ?

Exercice II. La durée de vie d'un système est une variable aléatoire X_1 de densité $f_\lambda(x_1) = e^{\lambda - x_1}\mathbb{1}_{x_1 \geqslant \lambda}$ (le paramètre $\lambda \geqslant 0$ étant la durée de vie minimale du système). On observe un n-échantillon $(X_1, \ldots, X_n)$, $n \geqslant 2$, de densité f_λ.

1. Trouver l'estimateur $\hat{\lambda}$ de λ par la méthode du maximum de vraisemblance et montrer que $\hat{\lambda}$ est une statistique exhaustive pour λ.
2. Calculer la loi de $\hat{\lambda}$ et étudier s'il s'agit d'une statistique complète. Peut-on construire un estimateur $\lambda^\star$ sans biais en ajoutant une constante à $\hat{\lambda}$? Calculer la variance de $\lambda^\star$.
3. Que vaut $\mathrm{E}(X_1)$? Utiliser la méthode des moments pour déduire un autre estimateur sans biais $\lambda^\sharp$ de λ. Calculer sa variance.
4. Comparer les estimateurs $\lambda^\star$ et $\lambda^\sharp$. Ces deux estimateurs sont-ils convergents ?
5. à l'aide de λ^* et d'une statistique complète, construire un estimateur efficace pour λ. Est-il unique ? Peut-on étudier son R-efficacité en utilisant l'inégalité de Cramer-Rao ?

6. Quelle est la loi de $\hat{\lambda} - \lambda$? En déduire que $2n(\hat{\lambda} - \lambda) \sim \chi^2(2)$. Indiquer deux valeurs $c_1 < c_2$ telles que $[\hat{\lambda} - \frac{c_2}{n}, \hat{\lambda} - \frac{c_1}{n}]$ soit un intervalle de confiance pour λ de coefficient de sécurité 95%.

7. Si $0 \leqslant a < b$ sont tels que $\mathrm{P}(a \leqslant \hat{\lambda} - \lambda \leqslant b) = 95\%$, construire un autre intervalle de confiance pour λ de coefficient de sécurité 95%. En déduire l'intervalle de confiance de longueur minimale. On pourra étudier la fonction $b = b(a)$ comme fonction de a.

8. Quelle est la limite en loi de $\sqrt{n}(\lambda^\sharp - \lambda)$, lorsque $n \to \infty$? Est-il possible d'en déduire un intervalle de confiance asymptotique pour λ de coefficient de sécurité 95% ?

Bibliographie

[1] Barra, J.-R., Baille, A.(1969) Problèmes de statistique mathématique, Dunod.

[2] Bickel, P.J., Doksum, K.A. (2001) Mathematical statistics : basic ideas and selected topics, Prentice-Hall.

[3] Borovkov, A.A. (1987) Statistique mathématique, Mir.

[4] Castell, F. (2002) Cours de Statistique pour l'Agrégation, Université de Aix-Marseille III.

[5] Casella, G., Berger, R.L. (1990) Statistical inference, Duxbury Press.

[6] DeGroot, M.H., Schervish, M.J. (2002) Probability and statistics, Addison-Wesley.

[7] Fourdrinier, D. (2002) Statistique inférentielle : cours et exercices corrigés, Dunod

[8] Garthwaite, P., Jolliffe, I., Jones, B. (2002) Statistical inference, Oxford University Press.

[9] Lehmann, E.L. (1959) Testing statistical hypotheses, John Wiley & sons.

[10] Lehmann, E.L. (1983) Theory of point estimation, John Wiley & sons.

[11] Milhaud, X. (2001) Statistique, Belin.

[12] Roussas, G.G. (1997) A course in mathematical statistics, Academic Press.